Reliability of Retrospective Event Histories within the German Generations and Gender Survey

Sauer, Lenore; Ruckdeschel, Kerstin; Naderi, Robert

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Bundesinstitut
für Bevölkerungsforschung

BiB Working Paper 1/2012

Reliability of retrospective event histories within the German Generations and Gender Survey

Lenore Sauer, Kerstin Ruckdeschel, Robert Naderi

URN: urn:nbn:de:bib-wp-2012-011

Preface

By comparing the Microcensus 2008 to the German GGS severe distortions were discovered by Michaela Kreyenfeld et al. (2010, 2011). This was reason enough for us to start a systematic analysis of the distortions and its possible causes together with Infratest, the polling institute which conducted the fieldwork of the survey. Our first aim was to correct the bias which was not successfully achieved due to the causes of the errors (see paper). Then, our second aim was to locate the distortions as accurately as possible in order to find solutions for further handling of the GGS and in order to provide valid data for users. We were in an intense discussion with Michaela Kreyenfeld which finally resulted in two parallel, but independent papers: The one of Michaela Kreyenfeld and her colleagues (2012 forthcoming) presenting the point of view of data users and our own paper focussing on the perspective and knowledge of the data producer and provider.

Abstract

Collecting retrospective data generally is accompanied by questions concerning the quality and reliability of such data. The aim of this article is to contribute to this body of research by analysing the fertility and partnership histories within the German Generations and Gender Survey (GGS). Although landmark events such as the birth of children or marriage are commonly regarded as reliable memories and therefore are assumed to be applicable retrospectively without much decrement in data quality, we find severe distortions in the retrospective data on fertility and partnership in the German GGS by comparison to the Microcensus. In fertility histories there is a major overestimation of childless women in older cohorts as well as an underestimation of this group in middle-aged cohorts. Regarding partnership history we have too many women never married in the older cohorts and too many married in the younger cohorts. We find that these distortions are mainly attributed to problems caused by survey organisation. The random route sampling procedure of the German GGS has led to problems, as has the instrument, which lacked control mechanisms while simultaneously being very complex. We also find indications of problematic interviewers, but we cannot delete the deviations in the data to our satisfaction as they are bound to be multicausal. We therefore conclude that there are combined effects of the sampling procedure, the complexity, and the length of the instrument that all contribute to interviewer effects.

1. Introduction*

There is wide agreement among social scientists that a better understanding of demographic behaviour – especially of family formation – is based on the life-course approach (for an overview on life-course research, see e.g. Elder Jr. 2009; Mayer 2009). Under this approach, one looks at family and fertility behaviour as processes that evolve interdependently of each other and of other processes in an individual's life course. Also cultural frames as well as institutional and structural settings play an important role in life-course research by relating micro-, macro- and meso-levels of analysis (Vikat et al. 2007: 394). There are two main ways to undertake life-course research. One is a prospective approach, which implicates a panel design of the survey and the inclusion of questions about expectations and intentions in the questionnaire. The other possibility is asking retrospective questions to cover a certain period of an individual's life (Blossfeld and Huinink 2001: 10f.).

The importance of collecting retrospective data that cover a long period of an individual's life is counterbalanced by questions about the quality and reliability of such data. Although demography has been notable in the past for its attention to data quality, there have been few recent evaluations of the quality of fertility or partnership histories. Although "landmark" events such as the birth of children or marriage are generally regarded as reliable memories and therefore are assumed to be applicable retrospectively without much decrement in data quality (Swicegood et al. 1984; Wu, Martin, and Long 2001; Hayford and Morgan 2008), more recent studies suggest that fertility information based on survey data might be subject to various types of distortions (Murphy 2009).

An inevitable difficulty in the validation of retrospective data is the availability of an objective source of information against which to compare survey responses. One possibility to validate retrospective information is the comparison with data that has been produced over a lifetime such as diary entries. If surveys containing identical retrospective questions are repeated, it is also possible to compare experiences of different cohorts (Babka v. Gostomski and Hartmann 1997: 131) or from the same individuals (Beckett et al. 2001). Another possibility is to compare reports from retrospective and panel observations on the same individuals (Peters 1988; Teitler, Reichman, and Koball 2006). If events in the life course generate a certain state such as being childless or being without a partner, then it is also possible to validate this information on an aggregate level with official statistics.

The first wave questionnaire of the Generations and Gender Survey (GGS) collects retrospective information on partnerships, fertility, the parental home, and home leaving. The possibilities for extensive validation of the fertility and union formation histories in the German GGS were quite limited due to the lack of comparable surveys or vital statistics. This changed in 2008 when for the first time the German Microcensus contained questions on childlessness and the total number of children of women (Statistisches Bundesamt 2009b; Pötzsch 2010). This now allows comparisons on this important part of the German GGS. Recent research on this basis shows various types of distortions, especially concerning fertility and partnership (Kreyenfeld et al. 2010; Kreyenfeld et al. 2011).

* We are grateful to Corinna Kröber who was a great help to us on the chapter concerning the state of the art on distortions in retrospective data. Also we would like to thank Ines Wickenheiser for providing us tables of the German Microcensus and Kai Dreschmitt for the layout of these figures. Also we thank Peter Ruckdeschel and Manfred Scharein for inspiring discussions about survey methodology, as well as Olga Pötzsch, Monika Pupeter, Heiko Rüger, Detlev Lück, Jasmin Passet, Michael Wagner, and the participants of the meeting of the German Panel Working Group for helpful comments. Furthermore we are grateful to Dora Kostova and Norbert N. Neuwirth for answering our questions on the implementation of the GGS in their countries.

Based on these initial research results this paper aims to verify the reliability of retrospective event histories concerning fertility and partnership within the first wave of the German GGS. Thereby the paper proceeds along the following structure: It first presents a summary of the state of the art concerning possible causes of these distortions (Chapter 2). Chapter 3 describes implementation of the GGS, and Chapter 4 is a detailed consideration of the main distortions in retrospective fertility and partnership history in the German GGS. To do so, the paper focuses not only on the extent of this bias, but also on identifying the most affected groups by making further differential analyses. The paper also provides possible reasons for these distortions given particular attention to the interaction between interviewer and respondent. In the article's conclusion, the results are used to formulate recommendations for further handling of the German GGS.

2. Previous research: Sources of possible distortions in retrospective data on fertility and partnership

Despite the crucial importance of the quality and reliability of retrospective data for demographic research, there have been few recent evaluations on this topic. The existing literature on errors in retrospective data can be summarised. First, existing literature highlights the importance of the characteristics or nature of the event itself (Chapter 2.1). Second, many empirical studies have investigated respondent characteristics associated with recall errors (Chapter 2.2). Third, previous research has revealed discrepancies that are attributable to survey organisation (Chapter 2.3).

2.1 Possible sources of distortions due to the characteristics of the event itself

Event history data requires responses on whether an event occurred and the timing of the event. Therefore the literature discusses the importance of the event as well as the passage of time between event and interview as possible causes for discrepancies. Another important aspect is the moment in which an event is remembered. Cognitive psychologists and survey methodologists attribute most flaws in data to cognitive operations involved in producing an answer (for an in-depth description of response problems see Sudman, Bradburn, and Schwarz 1996; Tourangeau, Rips, and Rasinski 2000).

The importance of an event is based on the concept of salience. Salience depends on many factors including pleasantness, emotional involvement, the event's role as a unique turning point in the lifetime, as well as its financial and social consequences (Wagenaar 1986). "Landmark" events such as the birth of children or marriage are commonly regarded as reliable memories and therefore are assumed to be applicable retrospectively without problems of distortions (Papastefanou and Tölke 1981; Swicegood et al. 1984; Klein and Fischer-Kerli 2000; Beckett et al. 2001; Wu, Martin, and Long 2001; Hill 2005; Hayford and Morgan 2008). However, more recent studies show that fertility information based on survey data might be critical and subject to decrement in data quality (Murphy 2009).

Another potential source of error in survey responses is the failure to remember relevant information. Human memories are fallible; as time passes, people omit dates and events, and sometimes they remember events only sketchily or inaccurately. Previous research has found that events in the distant past are underreported relative to recent events, with the degree of underreporting increasing as the time elapsed since the event increases (for a review of this literature, see Belli 1998; Wu, Martin, and Long

2001). Landmark events were thought to be exceptions to the general tendency for recall accuracy declining with time. However, more recent studies suggest that even dates of landmark events can be prone to error, e.g. for cohabiting or partnership relationships see (Klein and Fischer-Kerli 2000; Hayford and Morgan 2008).

It is useful to distinguish between event dates that are often "rehearsed" and those for which event dates are reconstructed by the respondent during the interview process. Events that are often rehearsed include landmark events such as the birth dates of one's children, marriage anniversaries and when a couple first met (Wu, Martin, and Long 2001). Rehearsal increases the ease with which memories are recalled, and failure to rehearse or recall a memory for a long time can make it difficult to retrieve it when it is required. It is then argued that respondents reconstruct data by employing simplifying strategies to construct answers during the interview. During this process events can be omitted through the inaccurate reconstruction of a memory (Sudman, Bradburn, and Schwarz 1996: 172ff.). If similar types of events take place, it then becomes more difficult to distinguish and recall specific events, what is referred to as "interference effect" (Tourangeau, Rips, and Rasinski 2000: 81ff.). Studies of recall accuracy in reporting of divorce dates (Mitchell 2010) or relationship timing (Reimondos, Evans, and Gray 2011) found some support for the interference hypothesis. A further finding from literature is the occurrence of "telescoping," which involves an incorrect timing of events. Often events are recalled as having occurred in the most recent period and fewer in the more distant past (Sudman, Bradburn, and Schwarz 1996: 186ff.) which leads to an overestimation of numbers in the recent reporting interval (Beckett et al. 2001: 596f.). A typical form of displacement is that births are moved forward, closer to the date of the survey resulting in a higher average age of mothers at birth (Singh 1987: 628f.). A further source of inaccuracy in memory is the addition of details over time. Most autobiographical memories consist of a multiplicity of information taken while or shortly after experienced. Information is added on later, when recounting the event to others or when simply thinking about it later on (Groves et al. 2009: 230).

In some cases respondents may resolve a reasonable response to the question, but choose to edit their true answers to survey questions either in order to portray themselves in a more favourable light and to avoid embarrassment or to give answers they think the interviewers will want to hear (Sudman and Bradburn 1974, 1980). One of the most common reasons for such a decision is the belief that certain behaviour or attitudes are socially desirable or undesirable (DeMaio 1984). Evidence of the occurrence of social desirability bias has recorded over-reports of participation in political elections (Traugott and Katosh 1979; Silver, Anderson and Abramson 1986; Holbrook, Green and Krosnick 2003; Karp and Brockington 2005); and under-reports of negatively associated behaviours such as smoking or bankruptcy. Similarly, other studies discuss the disparity between the reports by men and women about the number of opposite sex partners. They show that men tend to overestimate the number of partners they have had, while women tend to underestimate the number of partners (Tourangeau and Smith 1996, 1998). Other studies found that mothers revise their reports of whether or not they cohabited at the time of the birth of their child for fear of adverse repercussions (Teitler, Reichman, and Koball 2006). As negatively esteemed events or socially unacceptable behaviour will change over time, questions concerning the same events in different eras may produce different answers. For example, the increasing social acceptance of cohabitation in 2002 relative to 1988, for instance, may mean that cohabiting relationships that took place in the 1980s are more likely to be reported by respondents in the 2002 than in the 1988 survey (Hayford and Morgan 2008).

2.2 Possible sources of distortions associated to respondent characteristics

Furthermore respondent characteristics and attitudes may affect the quality of retrospective data. Particular attention is thereby given to the age of the respondents. Obviously the quality of answers decreases with age, because many memories are prone to lose specificity and detail with age. Having a child is generally assumed a reliable memory, because of all events that occur in a woman's lifetime, giving birth is one of the most memorable. But, evidence has yielded different results. Theoretically the number of children should increase with the age of the mothers, because in the past women gave birth to more children than women today. Analyses of the quality of fertility data in the World Fertility Survey show that older women are indicating a lower number of children than younger age cohorts (Singh 1987; United Nations 1987; Khan 2010). Empirical findings on the basis of the General Household Panel in Britain also show a negative effect of age on the quality of fertility data (Murphy 2009: 128ff.). Thereby it is assumed that an increasing level of inaccurate reporting of cohort childbearing has emerged with age among older women in recent times. Respondents at younger ages may have had less scope for providing misinformation since they would frequently have had children present in the household; this could have been increasingly less likely at older ages, making it easier to misreport information (Murphy 2009). It might be suggested that this is not an effect of age, but rather due to the circumstances of the survey organisation or situation (see Chapter 2.3).

Gender differences in the autobiographical recall of events are also frequently under study (Skowronski and Thompson 1990; Auriat 1991; Ross and Holmberg 1992; Skowronski et al. 1994). Klijzing and Cairns (2000) conclude on the basis of the Family and Fertility Survey (FFS) that women make better respondents than men, not only with regard to their willingness to disclose autobiographical information on life course events, but also with regard to their ability to remember the dates of these events. Most studies that include both female and male respondents reveal a gender effect with women being more accurate and reliable at dating the start or end of their relationship (Poulain, Riandey, and Firdion 1992; Klein and Fischer-Kerli 2000; Mitchell 2010).

Another variable that has been found to be related to the accuracy of recall is the education of the respondent. Empirical findings show that highly educated women are less likely to report inconsistent marital histories across waves of a survey (Peters 1988) and that highly educated respondents are less likely to misreport their divorce date (Mitchell 2010).

Other studies have found that highly educated women are more likely to give consistent answers concerning their desire for children and their methods of contraception (Coombs 1977; Smith and Thomas 2003). Thereby education is seen as a proxy of cognitive abilities. Cognitive abilities are also part of analysis from data of the US National Longitudinal Survey of Youth (NLSY) and the National Survey of Family Growth (NSFG) on the quality of answers concerning first sexual intercourse. Data quality varies significantly with duration of recall and measures of respondent ability related to arithmetic facility and memory (Wu, Martin, and Long 2001). Respondents who are motivated to devote the cognitive effort to search, retrieve, and integrate memories are more likely to provide high quality answers as are those who have a greater personal connection to the subject of the survey or those who believe that the survey will have useful consequences (Krosnick 1991).

2.3 Possible sources of distortions due to survey organisation

Finally, empirical findings show that the characteristics of the survey may also influence the quality of retrospective data. This kind of cause is associated with typical general survey problems such as distortions through the sampling procedure, the instrument, or the interview situation. Although these problems are not exclusive for event history data, special attention must be given to them in this context. Particularly the design of the questionnaire is a crucial factor for the accuracy of answers.

Empirical findings suggest that accessibility based on at-home patterns of different groups and the number of different persons in the household is one explanation for survey distortions (Groves et al. 2009: 193ff.). A different accessibility to mothers compared with childless women will lead to a (systematically) selected sample. Another example of different accessibility might be older childless age cohorts who are more likely to live in retirement homes, which are generally not included in sampling procedures of private households. In this context Murphy (2009) evaluates migration, mortality, and institutionalisation of living arrangements as possible confounders on the fertility history. He found only minor effects of these three factors. Another possible source of distortion is that married as well as cohabiting persons have a higher life expectancy. This could lead to a sample in which a larger number of older cohabiting persons were interviewed compared with singles of the same age group (Murphy 2000).

Some people may agree to participate in an interview as a result of a compliance process or because they are required to do so. After answering a first set of questions other respondents might become disinterested or distracted as the questionnaire progresses further (Krosnick and Presser 2010). For example, Murphy (2009: 130) suggests that over-reporting of childlessness could occur because some respondents may lose interest and report themselves as childless as a non-confrontational way of shortening the interview (even though at the outset, respondents are informed that they can discontinue the interview at any stage).

Also the instrument may be related to the accuracy of answers. The questionnaire can contain unclear wording or unclear terms of questions that could be interpreted distinctly, e.g. the term "cohabitation" (Knab and McLanahan 2007) or the term "children." The latter might be interpreted as young people regardless of their relationship to the respondent or the offspring of the respondent, regardless of their age. Some respondents may not know what a particular term means. The researchers who develop questionnaires are often experts about a subject, and they may overestimate how familiar the respondents are with the terminology they themselves use every day (Groves et al. 2009). Moreover, the reporting task a respondent confronts may be affected by the respondent's assessment of the characteristic being reported, which is usually tied to social factors (see also Chapter 2.2.). For example, a respondent with a complicated employment history will find it difficult to report beginning and ending dates of jobs, whereas this task will be simpler for someone who has held the same job since completing school (Schaeffer and Presser 2003: 68). A growing body of research suggests that survey results may be affected not only by wording of a question, but by the order in which questions are asked. It is known from panel or longitudinal studies that responses are influenced by previous experience in the same or similar surveys, which is known as survey or panel conditioning. Similar conditioning can occur within a single survey if the responses to survey items placed in the latter section of a questionnaire are affected by experience gained from earlier sections (Duan et al. 2007). Especially screening items that cause follow-up questions, which is typical of retrospective questions, are prone to this kind of conditioning. After learning that answering such questions in a certain way can lengthen the interview, respondents might answer incorrectly to subsequent screening items in order to avoid follow-up questions (Jensen, Watanabe and Richters 1999; Lucas et al. 1999; Duan et al. 2007; Kreuter et al. 2011). Accordingly survey responses in the latter section of a long questionnaire are biased towards underreporting.

A more common problem concerning the quality of retrospective data is excessive complexity, which means that the question has a structure that prevents the respondent from inferring its intended meaning (Groves et al. 2009: 228). Also a change of wording might influence comprehension of the question and therewith the answer (Swicegood et al. 1984; Murphy 2009). Respondents may also expand or restrict the meaning of concepts because the wording of a question evokes prototypes or exemplars that then dominate the definition of the concept (Schaeffer and Presser 2003). Also the ease with which respondents are able to recall events may also be influenced by the design of the questionnaire. An important aspect of the survey design is whether relationship dates are collected by a list of standard questions, or by context-based memory strategies such as an event-history calendar[1]. Event-history calendars were shown to significantly improve reporting about several variables – such as moves, income and weeks unemployed – although it increased over-reporting of other variables (Belli, Shay, and Stafford 2001; Belli et al. 2007).

The problem of wording is accompanied by the interaction between respondent and interviewer. When an interview begins, most respondents have little idea how they are expected to perform. It is then up to the interviewer to define what is expected of the respondent. Also, respondents who are uncertain about the intent of a question may ask the interviewer for clarification (see a detailed description of interviewer effect (Groves et al. 2009: 291ff.; Schaeffer, Dykema, and Maynard 2010: 450ff.). There is another body of research that has detected the influence of observable interviewer characteristics on respondent behaviour. For example both males and females report different gender-related attitudes to female interviewers than to males (Kane and Macaulay 1993). Other studies are focussing on the effect of the interviewer's age, education, or experience on the respondent's behaviour.

Another possible explanation is that if the interview takes place with other people present, such as a new partner, some persons may be reluctant to acknowledge earlier childbearing or partnership (Aquilino 1993).

3. Data

Obviously some of the distortions described below are related to the sampling procedure, the instrument, or its implementation. Therefore these aspects will be described in the following chapter in detail. Furthermore there will be a comparison of the GGS implementation in other countries in order to ascertain whether the national implementation may have caused odd results.

General description of the German GGS and its implementation

The target population of the German GGS is formed by German-speaking persons aged between 18 and 79 living in private households in Germany. This included all persons of the residential population of Germany who were linguistically able to follow the interview, regardless of their nationality or ethnic origin. The sample was taken on the basis of a random route survey according to the ADM design (ADM model as per 31 December 2002). This multi-stage sampling procedure comprises in a first step a regional stratification in so-

[1] The event-history calendar uses a central timeline from a predetermined starting point and respondents are asked to indicate the timing of various events across different domains of life, for example employment, education, residence, and family formation since the beginning of the timeline.

called sample points[2]. These sample points form the selection units of the first selection stage. In the second selection stage, the household addresses needed for the sample were collected by random walk selection. In each sample point a starting address is defined and using a precisely defined route, the interviewers are obliged to contact every third private household[3] and to carry out an interview if a respondent from the defined population has his/her primary place of residence there. If several persons live in the household belonging to the target population, the interviewer identifies the respondents who are to be asked with the aid of a Kish selection grid (Ruckdeschel et al. 2006: 11f.).

The survey was carried out in the period from 22 February to 12 May 2005. The contents were collected using face-to-face interviews based on CAPI software on laptops. A total of 528 interviewers were deployed. They were prepared by means of written instructions, which contained information on the selection procedure and on the realisation of the specifications of the interviews. It is notable that the average of roughly 19 interviews per interviewer varies largely from one interviewer to another. Nine interviewers tried in vain to complete any interviews, while 41 interviewers conducted 51 and more interviews (one interviewer carried out the maximum of 180 interviews). In total it was possible to carry out 10,017 analysable interviews, of which 7,760 were in the old and 2,257 in the new Federal states (Ruckdeschel et al. 2006: 20f.).

For the comparison of the characteristic distribution of the German GGS with reference statistics[4] we used the German Microcensus[5]. The German Microcensus is an official representative population sample containing structural population and labour market data. Its original intention was to provide updates between two population censuses. It is therefore carried out on a yearly basis in which one percent of all households in Germany are involved[6]. The content of the Microcensus and its questionnaire are regulated by federal law and participation is compulsory. The GGS shows the typical deviations of surveys (see in detail Table A 1 in appendix). Respondents from one-person households occur less frequently in the actual sample of the GGS compared with the overall population in the Microcensus (26 to 35%). This may be explained on both by the more difficult accessibility of economically active one-person households and by the higher willingness of persons with children to participate in the GGS (Festy and Prioux 2002; Hartmann and Schimpl-Neimanns 1992: 320). In other GGS countries this underrepresentation also occurs, albeit to a lesser extent (Statistik Austria 2009: 14). With regard to age distribution, it can be ascertained that large parts of the sample accurately portray the structures from the official statistics. Research shows that respondents with a higher level of schooling in particular tended to be more willing to participate in surveys (Hartmann and Schimpl-Neimanns 1992; Schnell 1997; Koch 1998; Neller 2005; Loosveldt, Carton, and Pickery 2008). A similar picture emerges for the GGS. Men of German nationality with a lower level

[2] For the first selection stage, the Federal Republic of Germany is sub-divided using the local, statistical districts and with the aid of a geographical information system (GIS) to sub-divide streets into roughly 53,000 areas. With the probabilities proportionate to the number of households in the sample points, the number of points was sampled in each cell: 1,173 of which were in western and 302 in eastern Germany.

[3] That means that residents of care institutions, penitentiaries, homes for the elderly and holiday homes are excluded from the sample frame, while students living in halls of residence are probably included.

[4] Such a comparison should take into account of whether effects exerted by different question and measurement concepts on the distribution of the characteristics can be largely ruled out with the characteristics in both data sources. As far as the results of the official statistics are consulted as a reference, this condition is only met as a rule for a small number of characteristics of the social structure.

[5] In the whole paper we used remote execution as form of access to the Microcensus. Remote execution is the only form of access permitting the analysis of formally anonymised original data. However, the user does not have direct access to the data. The data users receive structural data records (dummy files) which resemble the original material with regard to structure and the values of the variables. With the help of these dummy files, evaluation programs (syntax scripts) can be prepared using analysis programs, which will then be used by the statistical offices to analyse the original data. After the required confidentiality check has been made, the data users finally receive the results of that analysis (Statistisches Bundesamt, available online: http://www.forschungsdatenzentrum.de/en/datenzugang.asp, extracted on: 21-10-11).

[6] Approx. 370,000 households and 820,000 persons.

of educational attainment (*Haupt-/Volksschulabschluss*) are slightly underrepresented (38 to 42%), while German men with a high level of schooling (*Fachhochschule* or *Abitur*) are slightly overrepresented (29 to 27%). A similar picture emerges when taking a closer look at women of German nationality. Also in other GGS countries such as Austria (Statistik Austria 2009: 14) or France (Regnier-Loilier 2006: 19) persons with higher educational attainment tended to be more willing to participate in the survey. The composition of the population varies from GGS with a share of 5.5 percent of non-nationals to the Microcensus with a share of nearly 10 percent of foreigners. One possible reason is the focus on persons who are linguistically able to follow the interview, which might also explain the underrepresentation of lower educated men and women of foreign nationality. For example only 4 percent of the women of foreign nationality in the GGS state that they have left school without any certificate, while 20 percent in the Microcensus do so.

In order to avoid influences by these deviations in the following chapters we will concentrate on German nationals born between 1925 and 1987. As we especially consider fertility biographies we limit our analyses to German women who live in the western states of Germany (excluding Berlin). We do not consider the eastern states of Germany because the demographic situation in the eastern states is still very different from the one in the western states. Respondents are assigned as western Germans depending on the region they live in at the time of the interview. The German GGS (GGS_Wave1_Germany_V3.0) contains 10,017 respondents. After selecting females of German nationality of the cohorts 1925-87 who are resident in the western states of Germany, the sample size is reduced to 3,864 respondents.

Excursus: Description of GGS implementation in other participating countries

The GGS aims at international comparability by providing the survey design, common definitions, a standard questionnaire, and common instructions that each participating country should follow. Therefore it remains unclear why the distortions in fertility and partnership histories evolved only for Germany, while there does not seem such a bias for other GGS participating countries that used similar questionnaires (Kreyenfeld et al. 2010: 23).

A closer look at the information in Table 3.1.1 quickly reveals that an internationally perfectly comparable survey is not realisable. Data collection procedures as well as instrument and data editing differ slightly between countries, unfortunately resulting in major implications for fertility and partnership histories. Beginning with data collection one can see that the sampling procedure in every country is based on random selection. In most countries regional stratification is followed by register-based selection of households (France) or persons (Austria), while in Germany a random walk element is included. Therefore, part of the sampling procedure is in the responsibility of the interviewers and it becomes easier to deviate from prescribed interviewing procedures, such as conducting an interview with someone who is easily accessible and willing to participate in the place of the appropriate person. In Germany as well as the Netherlands these interviewers are part of a commercial field organisation, while France and Austria engaged interviewers from the national statistical offices. In all GGS countries the questionnaire was translated into the national language. Unfortunately in the German questionnaire an additional control question concerning the total number of children was not included, while with the exception of Bulgaria all other countries did so, although it was not included in the original questionnaire. Against this background it is obvious that different results concerning fertility and partnership might already be caused by the data collection, instrument, and subsequent data editing processes. By contrast to the other GGS countries Germany does not weight or correct the retrospective data on fertility and partnerships (see also the GGS guidelines for cleaning and harmonisation, Kveder and Galico)[7]. Due to all of these reasons, the analyses of reliability will be only examine the German example.

[7] Available online: http://www.ggp-i.org/materials/survey-instruments.html, extracted on 27-09-11.

	AUT	BGR	DEU	FRA	NDL	NOR
GGS as own survey or as part or in combination with another survey	Own survey	Own survey	Own survey	Own survey	As part of the Netherlands Kinship Panel Study (NKPS)	In combination with the 2nd wave of the Norwegian Life course, Ageing and Generation Survey (NorLAG)
Addition of administrative records	No	No	No	No	No	Yes, about 15% of questions were covered by administrative registry data
Target population	Resident non-institutionalised population aged 18-44	Resident non-institutionalised population aged 18-79	Resident non-institutionalised population aged 18-79	Resident non-institutionalised population aged 18-79	Resident non-institutionalised population aged 18-79	Resident non-institutionalised population aged 18-79
Sample size	5,000	12,858	10,017	10,079	9,765 of which 1,651 abbreviated version, 8,161 full-length	24,830
Sampling procedure	Stage 1: random selection of geographical units Stage 2: registry-based selection of persons	Stage 1: random selection of geographical units Stage 2: registry-based selection of persons	Stage 1: random selection of geographical units Stage 2: selection of households via random walk Stage 3: Kish selection grid	Stage 1: random selection of geographical units Stage 2: registry-based selection of households Stage 3: first name method	Stage 1: address sample based on Cendris (Dutch Mail) Stage 2: a) rules for dealing with multiple households at a single address, b) rules for dealing with different persons at a single household (next birthday method)	Sample stratified according to gender, age, geographical region and centrality of residential municipality
Mode of data collection	Face-to-face, computer-assisted (CAPI)	Face-to-face, computer-assisted (CAPI)	Face-to-face, computer-assisted (CAPI)	Face-to-face, computer-assisted (CAPI)	Face-to-face, computer-assisted (CAPI)	Computer-assisted telephone interviews (CATI) and self-administrated questionnaires on paper (SAQ)

Interviewer staff	Interviewer staff of Statistik Austria (National Statistical Office)	Interviewer staff of the Network of the Institute of Sociology at the Bulgarian Academy of Sciences	Commercial field organisation	Interviewer staff of INSEE (National Statistical Office)	Commercial field organisation	
Control questions concerning the total number of children	Yes, was automatically done during the interview	No	No	Yes, was automatically done during the interview	Yes, was automatically done during the interview	Registry data on births and children were added
Direct questions concerning marital status	Yes	No	No	No	Yes	Registry data on spouse and marital history were added
Weighting procedure	Step 1: sex, age, employment status, Step 2: country of birth, living arrangements (also including information on children), Step 3: cohort-specific parity distribution (only for the female sample)	No	Sex, age, education, region, number of household members	Sex, age, citizenship, social and occupational group, number of household members, type of household (including information on children) and size of urban unit	Step 1: address sample has to be converted into an individual sample, Step 2: sex, age, region, urbanisation, living arrangements (also including information on children)	Sex, age, geographical region, centrality and education
Imputations	Socio-demographic variables such as age, sex, education, and employment of the respondent and his/her partner and his/her parents in the household were corrected by imputations, as well as variables concerning income	No	No	Non-response was corrected by fitting		

Sources: Dykstra et al. 2005; Regnier-Loilier 2006; Ruckdeschel et al. 2006; Statistik Austria 2009; Buber 2010; Lappegård and Veenstra 2010; Bulgarian Academy of Sciences and Coordination Research Centre for Social Development and Social Eurointegration 2005, further information provided by Norbert Neuwirth (Austria), Dora Kostova (Bulgaria), own representation

4. Retrospective event history data in the GGS – problems and their explanation

The data about fertility biographies in the German GGS currently represent an important source concerning fertility in Germany and validation with reliable reference statistics is an important issue. Nevertheless possibilities of validation were limited before 2008, because up to then no official data about children ever born were available. In the Microcensus of 2008 for the first time women were asked about the total number of children they had ever given birth to. As the GGS provides a second source for validation of event history data with its partnership history, cross-validation is possible. This means that we can decide if distortions are due to special problems with fertility questions or if they are related to retrospective questions in general. We then can decide if distortions are caused by the sampling procedure, the instrument, or if they are inherent in the questions about fertility or partnership itself. In this chapter we therefore start with a comparison between the GGS and the Microcensus and show the results in Chapter 4.1.1. Thereby these results are presented in a descriptive way without offering any explanations for deviations with the exception of technical causes, such as the quality of the source of comparison and the different periods of data collection. The results will be explained in a separate chapter, followed by its explanations (4.1.2). In a second step we will proceed in a similar manner with regard to partnership history, by first showing the distortions (Chapter 4.2.1) and then by explaining them (Chapter 4.2.2).

4.1 Retrospective data on fertility

4.1.1 Description of distortions in fertility histories of the German GGS over cohorts

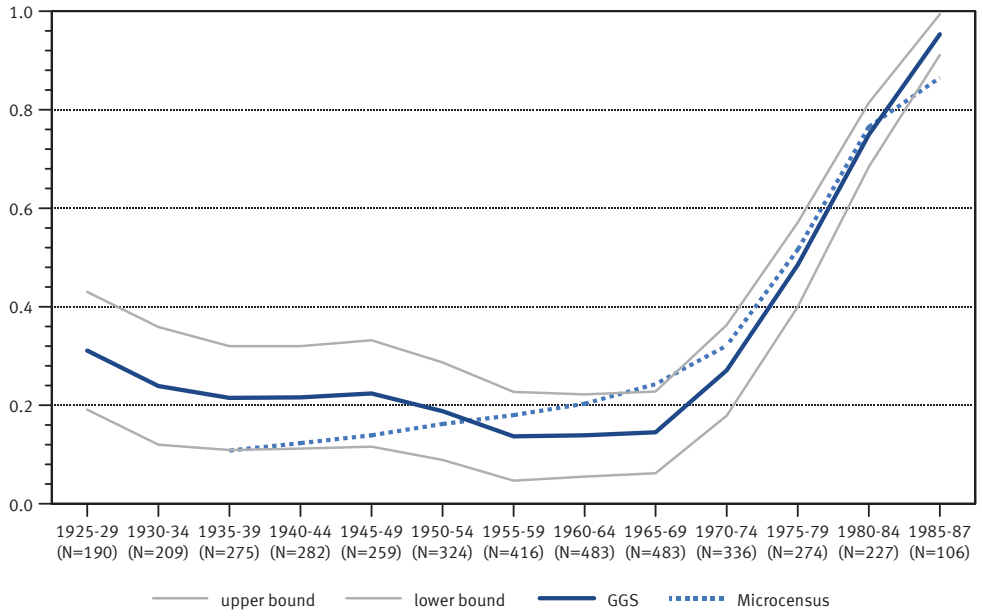
The German Microcensus 2008 offers an excellent opportunity for comparison and validation of fertility data. However, it should be noted that there are important differences between the GGS and the Microcensus concerning sample size and time the survey was conducted. There is a three-year gap between the Microcensus and the GGS and the sample size of the Microcensus, with about 820,000 persons, is much larger. The GGS is a random sample where the willingness of the respondents to participate is prerequisite while participation is compulsory for the Microcensus as part of official statistics (see Chapter 3). There is a duty of disclosure for nearly all questions and only a few, including the question about children ever born, are exempted from the obligation. The question about children ever born and their number were the only two questions in the Microcensus concerning this topic. This means that it is not possible to obtain further information on the timing and spacing of births.

By contrast, the GGS provides a wealth of data on fertility for both men and women. By collecting data on all household members, all biological, adopted, step, and foster children were recorded with information on sex and age. In another section of the survey information on all children living outside the household was gathered. In the so-called ‘child roster’ once again data on age, sex, and status, i.e. biological, adopted, foster, or stepchildren was recorded. With the information from these two parts of the survey the entire fertility history of a person can be reconstructed, including timing and spacing. The Microcensus does not allow any comparison of event data. The problems resulting from comparisons with other official and unofficial data have already been demonstrated elsewhere (see Kreyenfeld et al. 2010). Accordingly we do not investigate the subject of event analysis any further in this paper. The focus of our paper will therefore be on the proportion of childless women and on parity distribution.

Comparison: Proportion of childless women GGS 2005 – Microcensus 2008

The proportion of childless women across cohorts is presented in Figure 4.1.1. The curve shows an unexpected shape compared with the Microcensus because for the oldest cohorts till cohort 1960 the proportion of childless women is too high and there is a decline in childlessness (see Graph 4.1.1). Regarding the Microcensus as well as other sources childlessness should steadily increase over cohorts (Dorbritz and Ruckdeschel 2007: 50; Dorbritz 2008: 570). Direct comparison with the Microcensus shows two causes for this specific result of the GGS data. On the one hand there first seems to be an overestimation of childless women for cohort 1950 and older. Second, for cohorts 1955-65 there is an overestimation of mothers[8]. This is a conclusion also drawn by Kreyenfeld et al. (2011). On the other hand the results for cohorts 1970-83 are quite compatible with the Microcensus, showing only a slight overestimation of childless women in the unweighted data. This is a result of the fact that the GGS has an educational bias, i.e. more highly educated women are overrepresented in the data (see Chapter 3). In Germany, in these younger cohorts higher education means fewer children (Kreyenfeld and Konietzka 2008), which is reflected in the results. As education is included in the weighting factor this bias is corrected when using weighted data. With weighted data childlessness even seems to be overestimated in the GGS.

Figure 4.1.1: Proportion of childless women by cohort, Microcensus 2008 and GGS 2005



Notes: Women of German nationality living in western Germany
Microcensus cohorts starting with 1933
N=all members of selected cohorts in GGS
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

This is a logical result of the 3-year difference in data sampling. Age specific fertility rates for the respective cohorts were 82 children per 1,000 women aged 25 to 29 in 2006, for example, and 83 children in 2007[9]. These births lead to a lower proportion of childless women in 2008. For comparison and as a confirmation that higher age groups are less affected, age specific fertility rates for women aged 35 to 39 should be also mentioned. In 2006 they were as high as 42 children per 1,000 women and in 2007 it was 45 children (Statistisches Bundesamt 2011: 53).

[8] Although the Microcensus is always inside the bounds of the confidence interval, this cannot be treated as a confirmation of the GGS results, but merely does not allow rejection of the hypothesis that the two curves are different.

[9] Year of birth method (Statistisches Bundesamt 2011: 7)

The deviations of the GGS occur partly in cohorts with known problems in the Microcensus with regard to childlessness. We therefore need to check the extent to which the results of the GGS are concerned. To do this we must explain the problems of the Microcensus in more detail. We already mentioned that the response to the question on children ever born was not compulsory, only 89 percent of all women aged 15 to 75 answered it. The resulting missing values seem to be biased in a direction that affects our comparison. In the Microcensus the respondents could choose between a face-to-face interview, a telephone interview, and self-administered questionnaires. In the course of data preparation evidence has been found that there could be a systematic bias caused by the respondents who completed the questionnaire without an interviewer. It was them who generally did not answer the question about children ever born (item nonresponse). While interviews conducted by an interviewer showed less than 5 percent missing answers for this question the rates went up as high as 40 percent to 50 percent[10] for self-administered questionnaires (Statistisches Bundesamt 2009a: 1). From the compulsory questions we know that these women more often lived alone, were more frequently never married, and had a higher education than average. As they constitute a very particular group it can be assumed that answers concerning the question about children ever born are systematically biased. Probably childlessness is underestimated because women with that profile normally are more often childless than average (Statistisches Bundesamt 2009a: 2). To correct this bias a two-step imputation procedure[11] was implemented in order to minimise nonresponse as far as possible and to correct data on childlessness and number of children (for details see appendix). Still, a relatively high nonresponse rate remains after the imputation, especially in older cohorts[12], which has to be taken into account when comparing the Microcensus with the GGS. However, the results for the younger cohorts allow the assumption that in the Microcensus nonrespondent women of the cohorts 1950 and earlier are very often childless although nothing can be said about the exact numbers (see Statistisches Bundesamt 2009a; Pötzsch 2010 for this whole paragraph). This means the deviations in the GGS may be smaller than originally assumed.[13] The results of another survey about fertility in Germany from 2006 confirm this hypothesis albeit weakly (special survey "Births in Germany"[14] (Pötzsch and Emmerling 2008)). In this survey 14 percent of women of the cohorts 1931-41 as well as of the cohorts 1942-51 were childless (see Table 4.1.1). These results are more or less in the middle between the GGS and the Microcensus. They suggest that the overestimation of childless women especially in the older cohorts in the GGS may be a little less pronounced than a first glance would imply. Nevertheless this does not really change our results, because on the one hand both surveys do not reach the quality of the Microcensus and on the other hand also in this comparison the result is that childless women in the GGS are overrepresented in the cohorts of 1950 and older and that they are underrepresented in the cohorts of 1950-60.

[10] Depending on the age of the respondent

[11] First step: imputation of information on childlessness/motherhood; second step: imputation of information on number of children (see Statistisches Bundesamt 2009a for details).

[12] Beginning with the cohort of 1950

[13] A comparison of official birth statistics and of the Microcensus showed that average number of children in the cohorts 1938-68 in western Germany tends to be higher in the Microcensus than the CFR in the fertility statistics (Pötzsch 2010: 199). There is no way to determine if this indicates an underestimation of childlessness in the Microcensus. However, as the absolute deviation of the average number of children is never more than 0.03 for the older cohorts (Pötzsch 2010: 193) this should not affect our results.

[14] Sample: 12,458 women aged between 16 and 75; voluntary participation.

Table 4.1.1: Proportion of childless women by cohort in the GGS 2005, the special survey "Births in Germany" 2006 and the Microcensus 2008

Cohort*	GGS 2005 unweighted		GGS 2005 weighted		"Births in Germany" 2006		Microcensus 2008	
	%	N	%	N	%	N	%	N***
1931-41**	20.8	*132*	22.0	*145*	14	*1826*	10.0	*380*
1942-51	21.0	*117*	19.9	*115*	14	*1740*	13.2	*450*
1952-61	14.6	*124*	14.6	*107*	21	*1765*	17.4	*770*
1962-71	14.7	*147*	18.5	*155*	25	*2141*	22.5	*1,169*

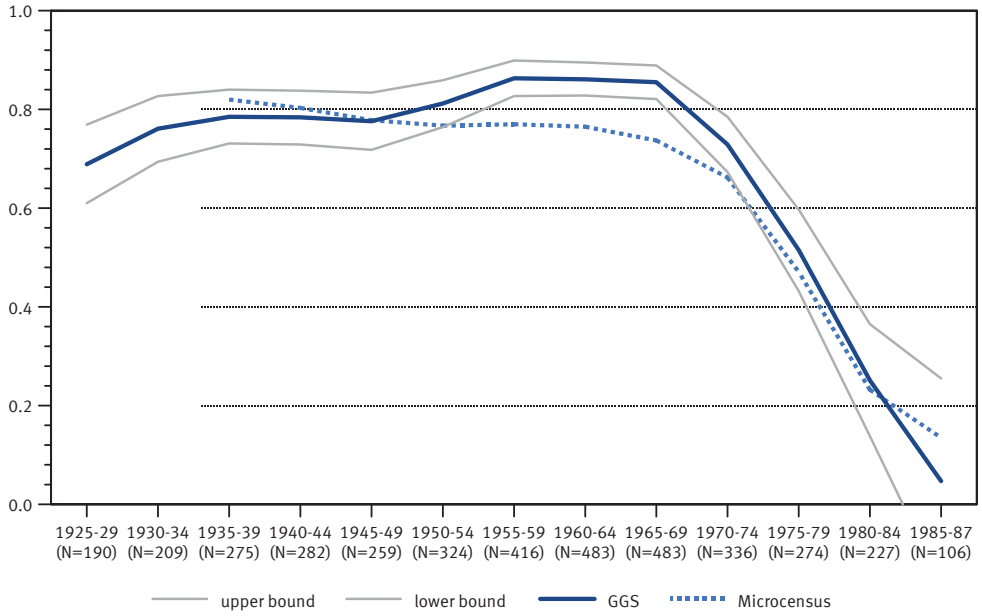
Notes: * Women living in western Germany without differentiation according to nationality
** Microcensus cohorts 2008 starting with 1933
*** Population in 1,000
Data sources: German GGS V 3.0 unweighted and weighted data; Microcensus by remote execution, own calculations

Comparison: Proportion of mothers, GGS 2005 – Microcensus 2008

Our next step was to compare the proportion of mothers. Normally this should simply be the reverse of the proportion of childless women. However, if our assumption is true that the missing values in the Microcensus 2008 for cohorts 1950 and older are frequently childless women, the proportion of mothers should be better represented when missing values are included in the calculations. Therefore we calculate in the Microcensus the proportion of mothers in relation to all women of a respective cohort including missing values[15]. In the GGS we cannot proceed in the same way, because we do not have any information about biases and their direction. In that respect the comparison of the Microcensus and the GGS indeed does not show any major deviations for the older cohorts (see Figure 4.1.2). However, for the cohorts 1950 54 to 1965-69 the proportion of mothers is considerably overestimated, which is in accordance with our results for childless women of these cohorts. The deviation of the GGS from the Microcensus is so large that it even falls outside the boundaries of the confidence interval. This means that the unusual shape of the childlessness curve of the GGS is not only caused by an overestimation of childlessness in the older cohorts, but in particular by an underestimation of childlessness in the middle aged cohorts.

[15] For the comparison of the proportion of childless women we compared all women who answered the question about children ever born, i.e. without item nonresponse.

Figure 4.1.2: Proportion of mothers by cohort, Microcensus 2008 and GGS 2005



Notes: Women of German nationality living in western Germany
Microcensus cohorts starting with 1933
Microcensus (including item nonresponse), GGS (not including item nonresponse)
N=all members of selected cohorts in GGS

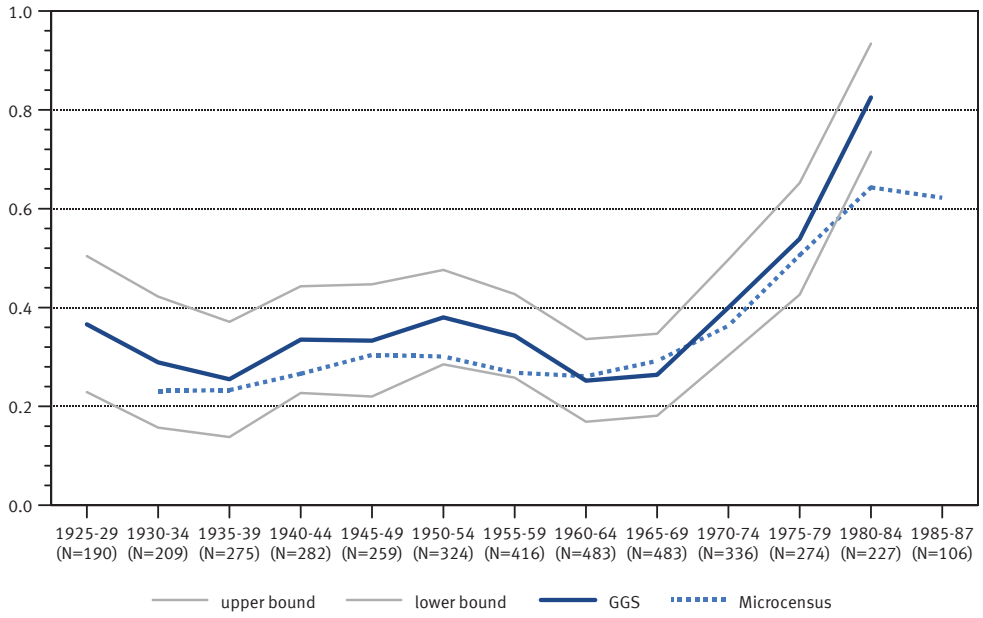
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

Comparison: Proportion of mothers by parity, GGS 2005 – Microcensus 2008

In the next section we want to focus on mothers only in order to make an internal validation of events that actually took place and in order to check if birth orders are correctly represented. More precisely, we want to compare parity proportions in the GGS and the Microcensus. For women with one child we find again two cohorts in which conformity is very low. For the younger cohorts 1975-87 the proportion of mothers with one child is too high and the proportion of mothers with two children is correspondingly too low (see Figure 4.1.3). One of the reasons is once again the gap of three years between the two surveys. According to age-specific fertility rates many of these younger mothers with one child in 2005 should have had another child three years later, which would also explain the high underrepresentation of parity two in the GGS for these cohorts (see Figure 4.1.3). As in the section concerning mothers in general, the middle-aged cohorts, i.e. 1950-59, show a peculiarity. Parity one is highly overrepresented here. The comparisons of parity three and parity four are quite satisfactory as there are no great deviations of the curves of GGS and Microcensus (see Figures A1 and A2 in the appendix).

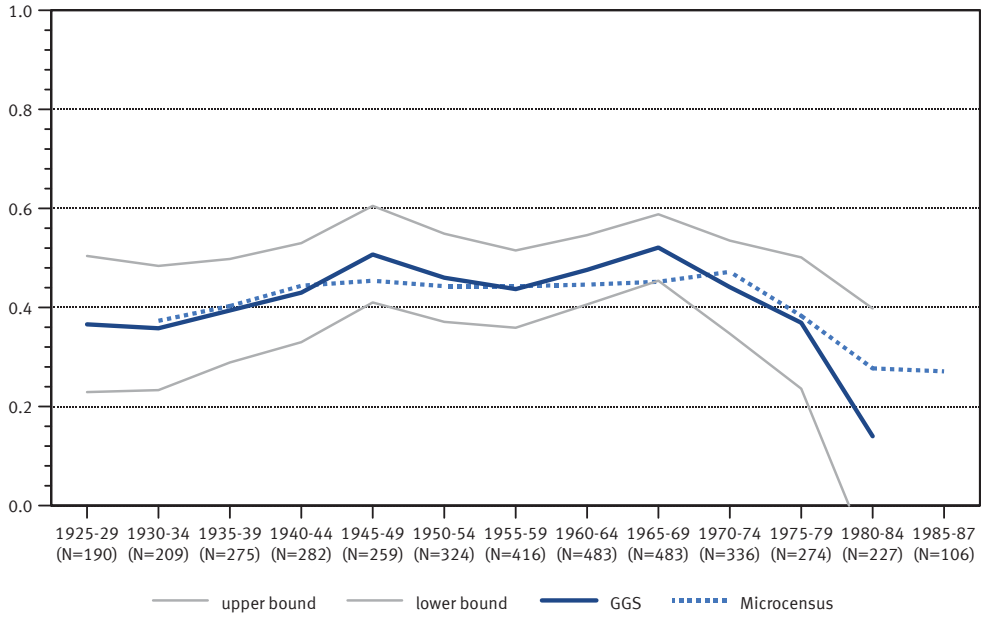
We can conclude this comparison with the statement that there are two significant differences between the GGS and the Microcensus: On the one hand we find an overrepresentation of childless women for the cohorts 1935-50 and an overrepresentation of mothers for the cohorts 1950-70. In the next chapter we will provide possible explanations for these distortions.

Figure 4.1.3: Proportion of mothers of parity 1 of all mothers by cohort, Microcensus 2008 and GGS 2005



Notes: Women of German nationality living in western Germany
Microcensus cohorts starting with 1933
N=all members of selected cohorts in GGS
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

Figure 4.1.4: Proportion of mothers of parity 2 of all mothers by cohort, Microcensus 2008 and GGS 2005



Notes: Women of German nationality living in western Germany
Microcensus cohorts starting with 1933
N=all members of selected cohorts in GGS
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

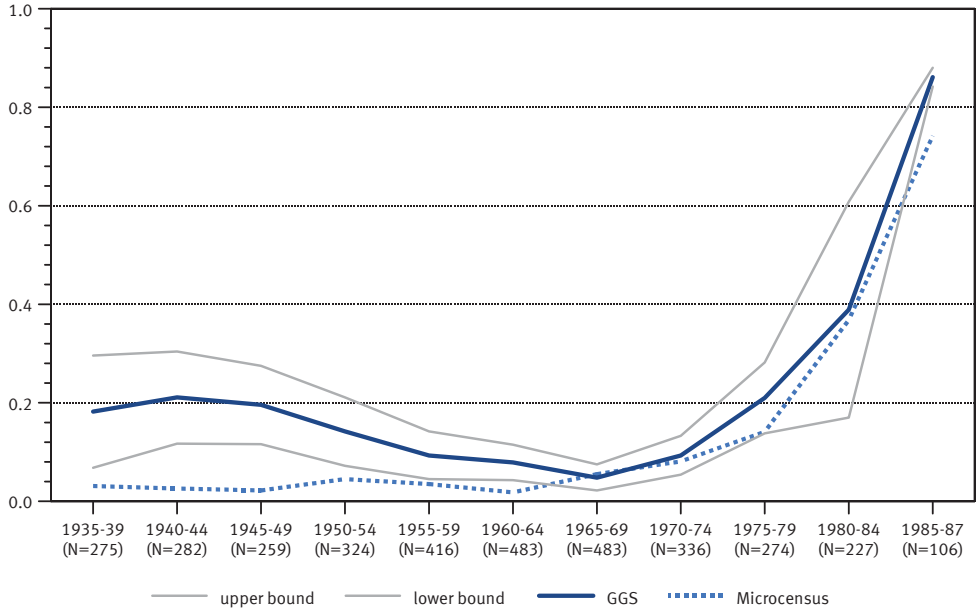
4.1.2 Possible explanations for the distortions in the fertility histories in the GGS

In the descriptive comparison of the GGS and the Microcensus some explanations for the differences between the two samples have been mentioned already. They are related mainly to peculiarities of the Microcensus and to the time gap between the two surveys. Nevertheless there are additional factors that explain the divergence between the curves of childless women and of parity distributions between GGS and Microcensus. In Chapter 2 we showed that there are different dimensions for explaining distortions in event history data. The first dimension was due to characteristics of the event itself. We can exclude this possibility here, because the birth of a child constitutes a very salient event and problems on this dimension are very unlikely (see Chapter 2, for a critical approach see Murphy 2009). This changes when we look for possible sources of distortions associated with respondent characteristics. We know that age and education in particular are factors that affect the quality of answers to retrospective questions (see Chapter 2.2). Older respondents more often give incorrect answers as well as respondents with lower education. Unfortunately we cannot re-contact respondents to verify their answers or have other means of crosschecking them. Therefore we cannot assign incorrect answers to individual respondents and confirm these hypotheses with the GGS. This leaves survey organisation as the last dimension we listed in Chapter 2 for possible sources of distortions. Here the focus is on the sampling procedure, the instrument, and the interviewer situation.

4.1.2.1 Sampling procedure and complexity of instrument

As for the sampling procedure, we are in the exceptional situation of having the possibility to compare the GGS to a second survey of 2006 with the same context, same questionnaire, and even the same social research institute that conducted the survey, but a different sampling procedure: the subsample of migrants of Turkish nationality living in Germany. This allows a perfect test of whether the sampling procedure is the cause for the distortions. For this survey a registry-based sample had been drawn, because for this subpopulation a random route technique is difficult to realise. The comparison of the proportion of childless women of Turkish nationality to the same population in the Microcensus 2008 reveals very ambivalent results: On the one hand, we find the same distortions as before for the older cohorts in which the proportion of childless women once again is overestimated. So we can exclude the sampling procedure for this part of the problem. On the other hand we do not find an overrepresentation of mothers in the middle-aged cohorts. This leads us to the well-known accessibility effect, which can often be found in random route samples (Esser 1974; Kohler 2007; Sodeur 2007). Especially in Germany mothers work less often than childless women making them easier to reach for interviewers. This concerns the accessibility of target households according to sampling rules as well as the availability of mothers when in fact other members in the household should have been interviewed (Koch 1998; Sodeur 2007). Therefore as a first result we can explain the overrepresentation of mothers in the middle age cohorts with the accessibility effect, indicating that interviewers perhaps did not take the sampling instructions as seriously as they should, which is a known problem with random route sampling (v. d. Heyde and Löffler 1993). In contrast, the overrepresentation of childless women in the older cohorts seems to be independent of the sampling procedure and the causes must lie elsewhere.

Figure 4.1.5: Proportion of childless women of Turkish nationality by cohort, Microcensus 2008 and GGS 2005

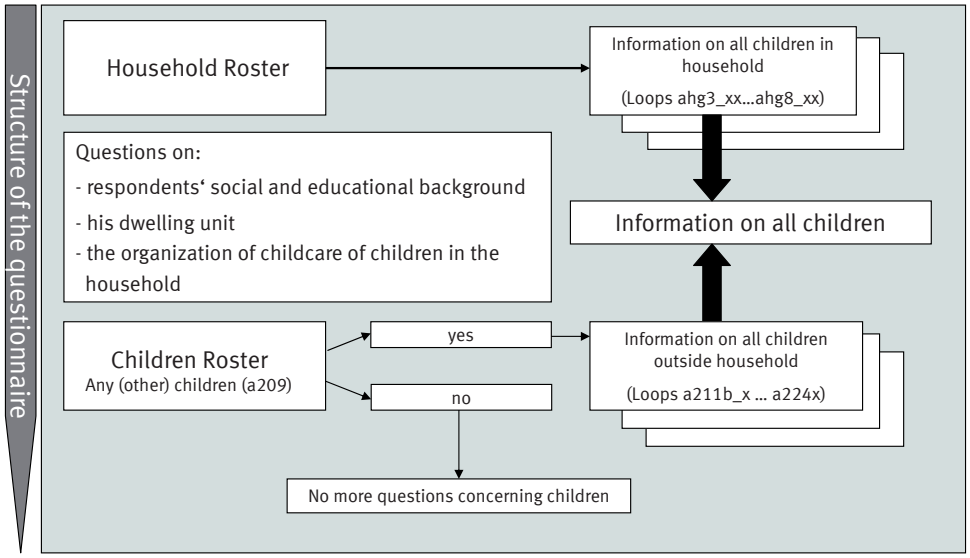


Notes: Women of Turkish nationality living in western Germany
Cohorts starting with 1935, because of a low number of cases
N=all members of selected cohorts in GGS

Data sources: German GGS, Turkish Subsample V 3.0 unweighted data; Microcensus by remote execution, own calculations

A second aspect of survey organisation is the instrument itself. The questionnaire of the first wave of the GGS was very complex in design and very demanding with regard to programming because of the many filters involved. Additionally the question about children was divided into two blocks about children in the household (included in the household roster) and children not living in the household, which were separate in the questionnaire (see Figure 4.1.6). In between the two blocks there were questions concerning the respondent's social and educational background, dwelling unit, and the organisation of childcare for children living in the respondent's household. This construction might have been too complex for respondents as well as for interviewers. We already mentioned that the survey designs differ quite substantially between the GGS countries in some respects, including the sampling of the information on children (see excursus in Chapter 3). To avoid mistakes control questions such as "Do you really have no children?" were implemented in Austria, in France and in the Netherlands, for example, but not in Germany or Bulgaria. To test the assumption that the construction of the survey is too complex, we constructed a variable that we called "memory gap." A memory gap exists if the question on children outside the household was positively answered, but there was no further information on these children in the child roster, such as status or year of birth. If the assumption of over-complexity is correct, countries that implemented control questions should have a lower memory gap than countries that did not implement such questions. The comparison of the extent of the missing information on children living outside the household (memory gap) confirms our expectations (see Table 4.1.2). The memory gap is considerably higher in Germany and Bulgaria than in the rest of the countries.

Figure 4.1.6 Location of the information on children in the GGS questionnaire



Own representation

Table 4.1.2: Extent of the "Memory Gap" in the GGS countries

	(Additional) children outside the household	Missing information on status of children		Missing information on year of birth of children	
	N	N	%	N	%
Bulgaria	2,146	149	6	241	11
Germany	1,891	96	5	101	5
Norway	3,293	69	2	1	-
Estonia	2,487	20	1	13	0.5
Georgia	1,913	13	0.5	0	-
Russia	3,073	10	1	12	1
Romania	2,752	4		15	
Austria*	180	2		1	-
France	2,364	0	-	2	-
Netherlands	1,785	0	-	0	-
Hungary	3,246	0	-	0	-

Notes: * women aged 18 to 44
Data sources: unweighted cases; GGS Austria, Bulgaria, France, Germany, Georgia, Hungary, Netherlands, Russia, Romania V3.0 respectively; Norway V3.1; women aged 18 to 79, own calculations

The complexity of the instrument might also lead to conditioning effects. During an interview respondents might learn that it might be beneficial to answer questions that are followed by loops with "no." Accordingly survey responses in the latter section, such as children outside the household, might be biased towards underreporting. This could also explain the overrepresentation of mothers with one child in the middle age groups. The cohorts of 1945-64 are exactly those cohorts where the children should start leaving the household or part of them should already have left. The overrepresentation of mothers

with one child therefore could be due to the fact that only the remaining children in the household were recorded whereas the children who already left were not accounted for. Once again this is only a suspicion, which we cannot really prove with the data at hand.

4.1.2.2 Interviewer effects

These previous results lead us to the topic of interviewer effects. In any survey in which data emerge from the interaction between interviewer and respondent there is a risk of falsification or cheating by interviewers, which could result in the contamination of the data. Several forms of falsification can be distinguished: 1) an interviewer fabricates all responses for an entire questionnaire, 2) an interviewer asks some questions in an interview and fabricates the responses to others, 3) an interviewer deliberately misreports disposition codes and falsifies process data (e.g. the recording of a refusal case as ineligible for the sample, 4) an interviewer deliberately miscodes the answer to a question to avoid follow-up questions, 5) an interviewer knowingly deviates from prescribed interviewing procedures, such as conducting an interview with someone who is easily accessible and willing to participate in the place of the appropriate person (Schreiner, Pennie, and Newbrough 1988; Groves et al. 2009: 319). Several factors may affect the prevalence of interviewer falsification: design factors relating mainly to questionnaire characteristics such as length, complexity, and difficult questions, organisational factors such as inadequate remuneration and training of the interviewers, as well as external factors such as bad weather or bad neighbourhood (Crespi 1945; Schräpler and Wagner 2003).

In the German GGS the interviewers who work part-time for TNS Infratest were paid per completed interview (regardless of the length of the interview) plus compensation for travel expenses. This could lead in few cases to the incentive to miscode the answers to questions in order to avoid follow-up questions and to save time. The mean value of the interviews' length was 57 minutes, but the duration of the interviews varied quite a lot. Short interviews took place above all in cases of people with a short partner biography in small household units (the mean value of interviews with 1-person households accounts for 52 minutes, for persons without former partner 49 minutes). Long interviews, by contrast, tend to signify large households in which the respondent had a longer partner history (the mean value of interviews with more than 2-person households accounts for 60 minutes, for persons with more than two former partners 63 minutes) (Ruckdeschel et al. 2006: 14).

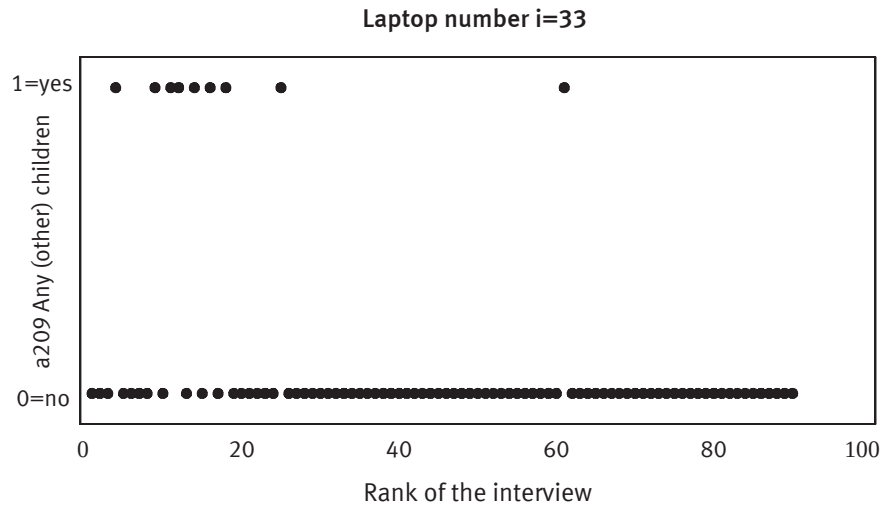
Procedures for detecting interviewer falsification include observation of the data collection process, re-contacting respondents, an ongoing review of data analysis concerning the effect of interviewers on the proportion of certain characteristics (Biemer and Stokes 1989) and actually, fraud detection using Benford's law (Schnell 1991; Schräpler and Wagner 2003). TNS Infratest used the verification method by re-contacting a sample of an interviewer's assignment in order to verify that an interview was conducted correctly (Ruckdeschel et al. 2006: 14). In addition we tried different methods to identify possible interviewer mistakes, because there is more than one possibility of tampering with interviews, which in turn requires different tests. As our findings can only reveal strong evidence, but no proof for falsification, we were quite conservative in these tests in order to avoid incorrect assignments.

On the one hand we evaluated interviewer effects in close relation to learning effects because of loops in event history data; on the other hand we evaluated them as a whole by taking a closer look at completed data records (see Chapter 4.2.2.2 below). For both approaches it was assumed that in at least partly fabricated interviews those answers dominate that lead to the avoidance of follow-up questions and therefore shorten the interview. This would have been quite easy in the GGS because of the structure of the questionnaire. If the answer to the question on children living outside the household was "no" and there were no control questions, the interview could have been reasonably

shortened (see Figure 4.1.6 above). Very similar effects are known from research on interviewer-related variation in network size (Van Tilburg 1998; Marsden 2003; Wolf 2004). Wolf (2004) compares the effect of interviewer-related variation on network size in different surveys. One explanation of this variation is that of differently experienced interviewers. More experienced interviewers know how to shorten interviews in order to elicit fewer friends. Those experienced interviewers generally work at commercial market research institutes. In surveys carried out by such commercial institutes the network size is smaller than in self-administrated surveys with less experienced interviewers (Wolf 2004: 265ff.).

In our first approach of interviewer control we therefore assumed that there could be a learning or an experience effect: with increasing number of interviews the interviewers learned how to shorten interviews to their advantage. To test this hypothesis we used a simple logistic regression approach by taking the answer to the question if there were children outside the household as the dichotomous dependent variable. We used the same approach for learning effects concerning partnership (see Chapter 4.2.2.2 below). We applied the rank[16] of the interview as independent variable. Additionally we controlled for age of the respondent, because the fact of having children who already left home is strongly dependent on the age of the child for which the age of the respondent is a good proxy. We did not exclude male respondents for this analysis, as we did not expect any gender effect here while at the same time it was important to get information on all interviews of an interviewer. Assuming that a large number of interviewers only conducted a very small number of interviews, which on the one hand could lead to extreme parameter values and on the other hand should have only a small effect on total results, we concentrated in our interviewer monitoring on interviewers with ten and more interviews. This left us 275 interviewers[17], who conducted 8,861 interviews. When taking into account all of those 8,861 interviews the results of the logistic regression analysis showed no significant effect on rank of the interview. As expected, age had a significant effect in that with greater respondent age the probability of having children outside the household increased. We then repeated this analysis separately for each interviewer and checked for significant effects up to the 5 percent level. As a result we identified eight interviewers who completed 468 interviews in total where the rank of the interview had a significant effect on the probability of having a child outside the household.

Figure 4.1.7: Example for a significant negative learning effect of one interviewer



Own representation

[16] It selects the interviews assigned to one interviewer according to the time when an interview was carried out, i.e. rank 1 symbolises the first interview, rank 2 the second one and so on.

[17] In fact we have information on laptop numbers. 519 interviewers used 524 laptops. So, we assume that laptop and interviewer number are identical.

Of those eight interviewers four showed a significant negative effect, which supports our assumption that they learned how to shorten interviews (see also Figure 4.1.7). The other four had a significant positive effect, which supports the idea of unskilled interviewers who learn how to handle the questionnaire better with each interview. Although these are opposite effects the result for the quality of the data is the same: negative. Nevertheless the number of detected interviewers is quite small and far from proving of our assumptions, as they also could be random results. The interviews constitute about 5 percent of all interviews and therefore the effect of omitting them cannot be too great either. We tested it nonetheless and found slight reductions of about 1 to 2 percent points of childlessness in the cohorts 1925-40. We finally checked the interviewers with no positive answers at all on the question on children outside the household separately. For that reason we were looking for interviewers with no positive answers on the question about children outside the household and a share of 50 percent of respondents aged 60 or higher. We opted for the benchmark of age 60 to make sure we only had respondents whose children already should have left home. We did not find any problematic interviewers, even if we allowed for a maximum of 10 percent positive answers to the respective question.

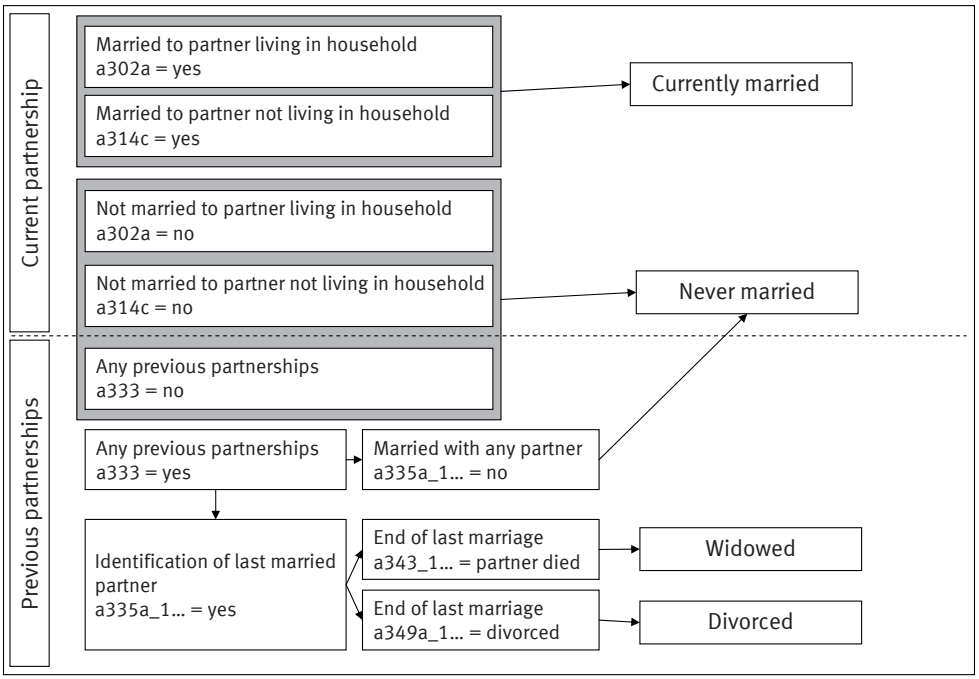
As a first conclusion on the causes of the distortions in fertility history data in the GGS we find effects of the sampling procedure, the complexity, and length of the instrument. All these effects add to interviewer effects in the sense that random route allows a great deal of uncontrolled interviewer actions as well as the instrument lacks control mechanisms while simultaneously being very complex. We found problematic interviewers, but we could not delete the deviations in the data to our satisfaction as they are bound to be multidimensional.

4.2 Retrospective data on union formation history

With the history on union formation the GGS provides a second source for validation of our findings. If there are special problems associated to retrospective data, the history on union formation or partnership history must lead to similar results as the fertility history (for an analogous proceeding see Kreyenfeld et al. 2010).

As we aim to validate the retrospective section of the survey on partnerships with official statistics we must limit our analyses to marital status, because it is the only available comparable indicator in official statistics. Marital status consists of four categories: married, never married, divorced, and widowed. Within the GGS it can only be identified by the information given in the loops about previous unions, because there are no direct questions concerning marital status (see Figure 4.2.1). Information about current marriages is asked in a special chapter of the questionnaire about the current partnership. The category never married can only be constructed by combining the union history with the current status. A respondent can be identified as never married if he has no current and had no former marriage. The marital statuses widowed and divorced can be constructed using variables from the section dealing with former relationships. Respondents who are presently not married were asked if they had former partners and if they affirmed, they were asked how that relationship ended. The marital status then can be calculated by using the information on the end of the last marriage if there was any. Additionally the dates of all former partnerships are saved in that section of the survey. Therefore the proportions of those four martial statuses can be compared and validated with Microcensus 2005 and 2008 for example in the chapters concerning fertility. The Microcensus 2005 will be applied if we focus on marital status only. In a later part of this chapter we will include parenthood – this is when we will switch to the Microcensus 2008.

Figure 4.2.1: Construction of marital status using the retrospective questions on partnership history



Own representation

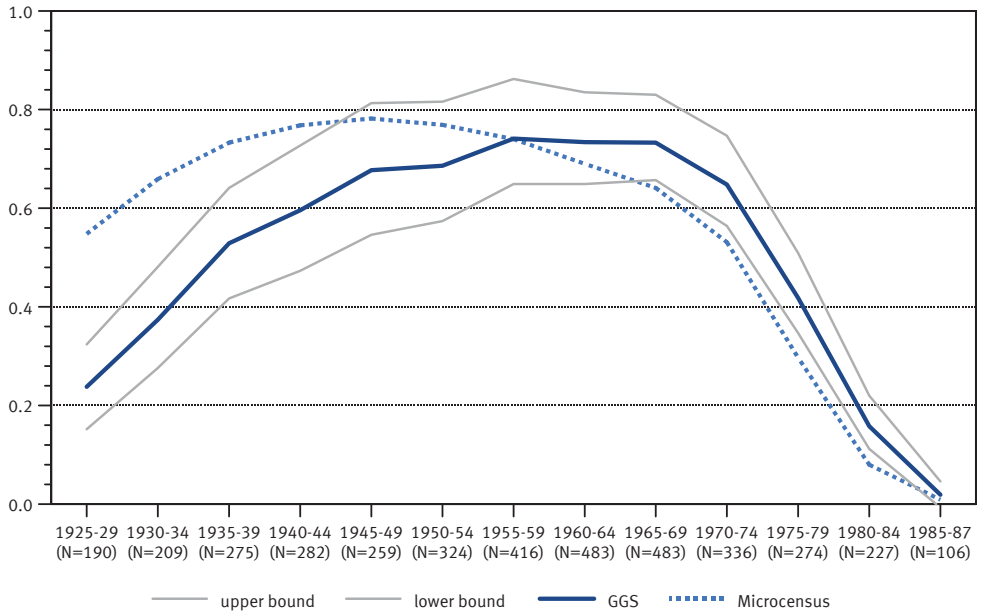
4.2.1 Description of distortions in union histories of the German GGS over cohorts

Comparison: Proportion of married women, GGS 2005 – Microcensus 2005

Analogously to the previous chapter we will begin by describing the comparison of the GGS and Microcensus on the proportion of married women across cohorts as presented in Figure 4.2.2. Especially in the cohorts 1925-44 there is a huge gap between the curves of both data sources, with the GGS being 30 percent points lower than the Microcensus for the oldest cohorts. The lines get closer for the cohorts of 1945-64. After the cohorts of 1965 the difference grows again and is outside the confidence bounds, but never reaches more than 13 percent points deviation. This means only four cohort groups (1945-64) are within the confidence bounds when looking at the marital status "married," although the data is better for younger cohorts than for the older ones. Thus the results for the GGS differ substantially from the Microcensus, but even without this comparison they would be peculiar. The curve in Figure 4.2.2 represents a picture of two overlapping effects, a cohort effect, and an age effect. On the one hand the cohort effect should show a more or less continuous decline in the proportion of married women, because marriages rates in western Germany declined for most of the last century, with some exceptions caused by World War I and II and the global economic crisis in the 1920s. From the 1960s onwards, i.e. for cohorts 1945 and younger the marriage rates decrease continuously (Dorbritz 2008; Winkler-Dworak and Engelhardt 2004, Bundesinstitut für Bevölkerungsforschung [18]). On the other hand more and more of the older married women become widowed, which indicates an age effect overlapping the cohort effect. This is reflected quite nicely by the curve of the Microcensus, but not in the GGS. In this dataset the proportion of married women is too low in the older cohorts, even if considering widowhood. Furthermore the decline of the proportion of married women begins 20 years too late in the GGS, i.e. only with cohorts 1965-69.

[18] Available online: http://www.bib-demografie.de/cln_090/nn_1758206/DE/DatenundBefunde/04/Abbildungen/a__04__01__eheschl__d__1841__2009.html, extracted on 02-10-2011.

Figure 4.2.2: Proportion of married women by cohort, Microcensus 2005 and GGS 2005

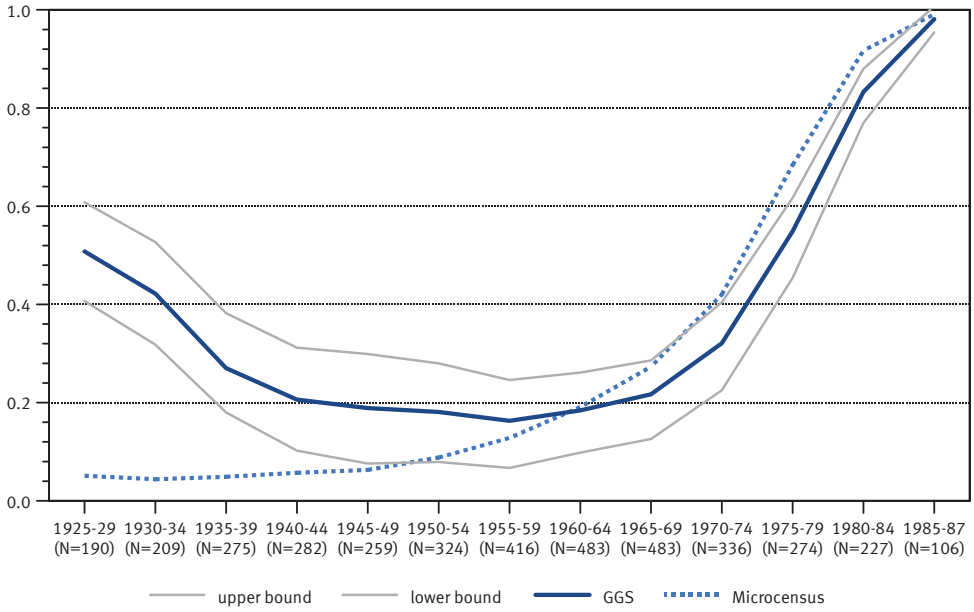


Notes: Women of German nationality living in western Germany
N=all members of selected cohorts in GGS
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

Comparison: Proportion of women who never married, GGS 2005 – Microcensus 2005

By taking a closer look at women who never married one should expect nearly the reverse of the proportion of married women. This holds true: While the proportion of married women in the older cohorts was too low, the proportion of women who never married is far too high (about 45 percent points in the cohort of 1925-29 and 38 percent points in the cohort 1930-34 (see Figure 4.2.3). Such a high percentage of women who never married is simply unrealistic. As we already argued when discussing the proportion of married women (see above) the proportion of women who never married should decrease with higher age and older cohorts far below 10 percent, while the number of widowed women should increase (see Brüderl and Klein 2003: 197; Backes and Clemens 1998: 44). As with the results on married women the proportion of women who never married in the younger cohorts, beginning with 1965 is too low, although the deviations are far from being as strong as in the older cohorts.

Figure 4.2.3: Proportion of women who never married by cohort, Microcensus 2005 and GGS 2005



Notes: Women of German nationality living in western Germany
N=all members of selected cohorts in GGS
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

4.2.2 Possible explanations for the distortions in the union formation history of the GGS

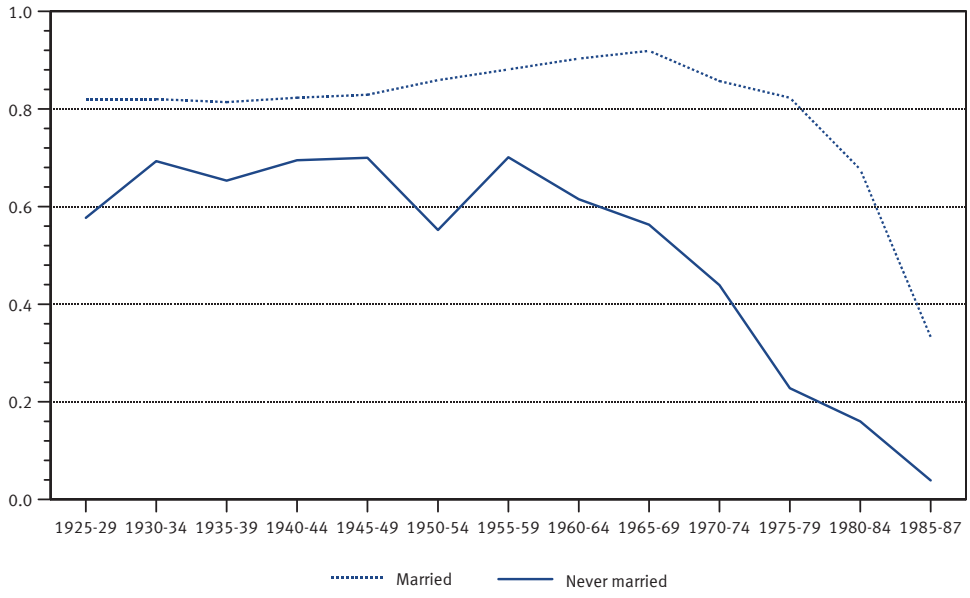
As in the chapter on the distortions in fertility history we will concentrate on the survey organisation as a dimension for explaining possible distortions in event history data. As only partnerships with a common household were subject of the survey, it can be once again assumed that the salience of the event is very high, so that this might not be a possible source of distortion. Instead there could be a sort of avoidance, i.e. avoidance by interviewers or by respondents because of embarrassment over intimate questions or because of conditioning effects keeping the interview shorter. The first should be especially pronounced if third persons were present at the interview. We tested this hypothesis, but these third persons often were married partners, which leads to some kind of tautology. Having a married partner present at the interview will lead to good information on marital status. So we did not find any effect for the category "other persons." Concerning respondent characteristics we cannot check the answers on individual level. So, once again survey organisation remains to be checked. We will proceed analogously to Chapter 4.1.2.

4.2.2.1 Sampling procedure and complexity of the instrument

The sampling procedure, which we already detected as the cause for the distortions in the middle-aged cohorts in fertility history, can also be related to the oversampling of married women in the younger cohorts. In western Germany the link between marriage and childbirth is still quite strong and therefore married women more often have children than women who never married (Dorbritz 2008: 573). This also implies that they work less often. As a result they are more easily accessible to interviewers (see Chapter 4.1.1).

To validate this assumption with our data we checked the proportion of mothers for married women and women who never married. Figure 4.2.4 shows that our assumption holds true. Especially in the middle aged and younger cohorts among whom we detected the accessibility problem, the proportion of mothers is significantly higher for married women (of all married women) than for women who never married (of all women who never married). Nevertheless a different and even greater problem occurs when looking at Figure 4.2.4. The proportion of mothers among women who never married is much too high for the older and the middle-aged cohorts[19]. This finding contradicts all known facts about the link between marriage and birth in western Germany (Dorbritz 2008; Kreyenfeld, Konietzka, and Walke 2011: 171). The results for women who never married should be the other way round, i.e. very low for the older cohorts and slightly higher for the younger ones (Dorbritz 2008: 573; Bundesinstitut für Bevölkerungsforschung[20]).

Figure 4.2.4: Proportion of mothers among all married and among all women who never married by cohort, GGS 2005



Notes: Women of German nationality living in western Germany
Data sources: German GGS V 3.0 unweighted data; own calculations

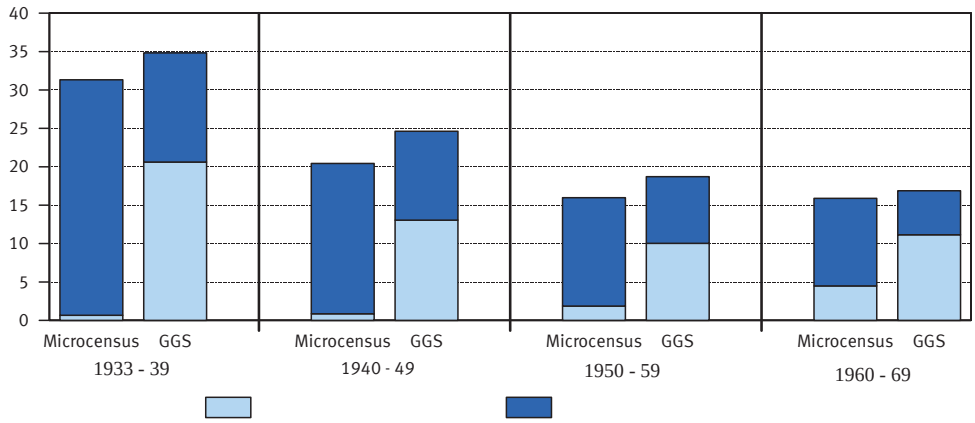
As with the results on distortions in fertility history we once again checked the complexity of the instrument for a possible explanation of the high proportion of mothers who never married in the older cohorts. It was quite easy for interviewers to avoid the loops on previous partnerships by simply indicating "no" on the question if any previous partnership existed, because no control questions were implemented (see Figure 4.2.1 above). To test this assumption we once again looked at the so called "memory gap" (see Chapter 4.1.2.2), which in this case indicated that women answered the question on previous partnerships with "yes" but did not give any further information on these partnerships. In total, out of 995 women who indicated a previous partnership 90 women, or about 9 percent, did not provide any further information. We take this as another indicator for the over-complexity of the instrument and the importance of control questions in the questionnaire. Additionally avoiding the loops on previous partnerships would be especially tempting for mothers, because they already experienced the loops on children earlier in the survey. If our assumption about some kind of conditioning is

[19] Nevertheless we must bear in mind that we are speaking of relatively low actual numbers, see table A2 in the appendix.

[20] Available online: http://www.bib-demografie.de/cln_099/nn_1881710/DE/DatenundBefunde/06/Abbildungen/a__06__03a__ehel__nichtehel__lbdgeb__w__1946__2009.html, extracted on: 24-08-2011.

true most of the women who never married in older cohorts should be widowed while the ones in the middle aged cohorts should be divorced. We tested this assumption by calculating the sum of all women who were widowed, divorced, and mothers who never married because the sums in the GGS and the Microcensus should be more or less equal, if widowed and divorced mothers are simply substituted by "never married" mothers in the GGS.

Figure 4.2.5: Proportion of never married or widowed / divorced mothers, Microcensus 2008 and GGS 2005



Notes: Women of German nationality living in western Germany
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

Our assumption goes along with the results presented in Figure 4.2.5. There the differences between the sums of never married, divorced, and widowed mothers never exceed 4 percent, with the largest difference in cohorts 1940-49. This can easily be attributed to random effects that occur in every survey. However, if we take a closer look at the different marital statuses the differences are enormous and cannot be attributed as random anymore. While we find less than 1 percent of mothers who never married in the Microcensus for the cohorts 1933-39 this proportion is about 20 percent in the GGS. These differences diminish across the cohorts: from 20 percent for the oldest cohorts, over 12 percent for cohorts 1940-49, to 8 percent and 7 percent for the two younger cohorts respectively. The picture is reversed for the proportion of widowed and divorced mothers. These findings support our assumption of the substitution of widowed and divorced mothers with mothers who never married. It is not possible to differentiate between miscoded and real divorced/widowed mothers, so we were not able to distinguish between them. All one can assume is that divorce rates were higher in the middle-aged cohorts (Dorbritz 2008). There should be a higher proportion of divorced mothers while in the older cohorts widowhood should prevail. Once again it is very likely that either respondents or interviewer avoided starting the loop on partnership.

4.2.2.2 Interviewer effects

These last results again lead to the topic of interviewer effects as we are dealing with a very complex section of the questionnaire where control questions are lacking. As in the section of the questionnaire on fertility history, interviews could be considerably shortened by avoiding the loops on previous partnerships. Therefore we checked for learning effects of interviewers once again. Our assumption was that with increasing rank of interview the probability of a positive answer on the question of previous partnerships should get smaller. Again we used a simple logistic regression model to

test it with previous partnerships as the dependent variable and rank of interview as the explanatory variable. By contrast to the fertility history no clear influence of age can be modelled in partnership history, because on the one hand the probability of having had a previous partnership should increase with age and on the other hand social norms in Germany were such that in older cohorts the number of previous partnerships should be quite limited (Schmidt et al. 2006: 27f.).

The results of the logistic regression revealed no significant effect when all 275 interviewers are included. Hence, a general contamination of data by interviewers can be excluded. However, when we tested for interviewers separately we identified thirteen interviewers with a significant effect on the answer to the question about previous partnerships. Six out of those thirteen interviewers had a significant negative effect, which is consistent with our assumption. The other seven interviewers showed significant positive effects, which again indicates a possible learning effect in terms of better handling the questionnaire. These thirteen interviewers together conducted 907 interviews. Compared with the interviewer control on fertility history two aspects must be mentioned. First, there are more conspicuous interviewers here than in fertility history, which can be explained by the fact that partnership history is a later section in the questionnaire than fertility history. There could be a tiring or conditioning effect during the interview and more interviewers omit this second block of loops. Secondly, the interviewers who are conspicuous in the section on fertility history are not the same as the ones in partnership history. There are two explanations for this fact. On the one hand interviewers are not always falsifying systematically, but more or less erratically and we only used a very conservative and simple test to detect them. On the other hand this does not have to be only an interviewer effect but could be caused by respondents, as well. Unlike the interviewer who has a growing experience of the structure of the questionnaire with each additional interview, respondents only have the chance to learn during the one interview they are actually performing. This means they would answer questions on children outside the household but might be hesitant to answer questions on previous partnerships because now they know that this might be time consuming. The proportion of mothers who never married in the GGS might serve as a sort of confirmation for this assumption (see Chapter 4.2.2.1). However, regardless of the actual cause the effect on the quality of data is negative once again.

Both interviewer controls share the assumption that with the progression of time interviewers learn how to shorten interviews. But, some interviewers might be very experienced, because they already conducted other large surveys and they did not learn during the GGS, but before. This led us to perform a more general control of interviewers. For this approach we checked not only extreme parameter values of interviewers, but also length of the interview and time intervals between interviews to detect possible irregularities. Once again we concentrated on interviewers with ten and more interviews. Furthermore it was assumed that interviewers who fabricate all or some responses in the questionnaire would indicate that the respondents were not willing to be re-contacted. The mean value of willingness to be re-contacted was 64 percent. So, all interviewers with more than ten interviews and a mean value of willingness to be re-contacted of below 60 percent were included in our interviewer control. This concerns 104 of originally 519 interviewers, who conducted nearly 30 percent of all interviews (2,693 interviews).

In 32 cases the willingness to be re-contacted was zero and in further 17 cases ten percent and less (see appendix Table A2). In one case 130 interviews were conducted without any respondent who agreed to be re-contacted, and in 1,390 conducted interviews only 30 respondents were willing to participate in a second wave! All designs for estimating the effects of interviewers on proportions of certain characteristics share the feature that different interviews are assigned to equivalent respondent groups. Using this equivalency, differences in results obtained by different interviewers are then attributed to effects of the interviewers themselves. Then we analysed if there were peculiarities in the distribution of answers concerning parenthood, the present partnership, and the

number of previous partnerships. Special focus was directed at the share of childless respondents, the share of persons without a present partner, and those persons without any or only one former partner.

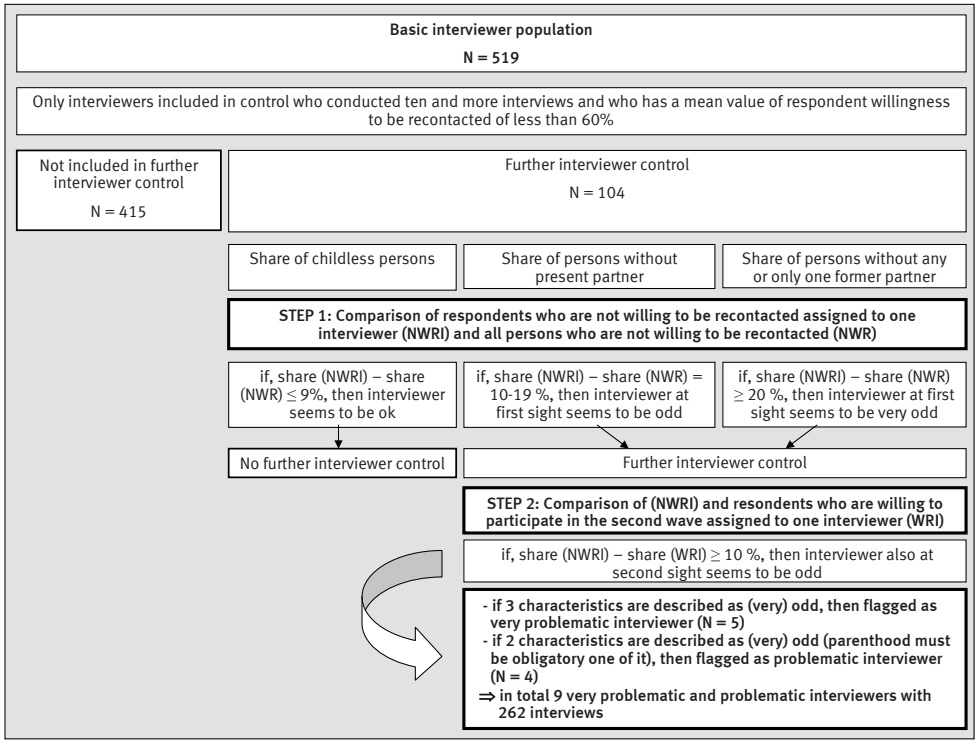
Concerning these three (constructed) variables in a first step we compare the difference between the proportion of a characteristic of those respondents who are not willing to be re-contacted assigned to one interviewer (NWRI) and the proportion of this characteristic of all respondents who are not willing to be re-contacted (NRW) (see Figure 4.2.5 below). Therefore we only take account of those interviewers if the difference between the proportions results in less time and effort: For example 37 or 39 percent of NRW indicate that they are childless or without a present partner. If the share of NWRI is lower, then this interviewer is not further monitored, if this share is higher, then this interviewer is under further control. Every interviewer for whom this difference exceeds 20 percent is flagged at first sight as very odd and every interviewer where this difference is between 10 and 19 percent is flagged at first sight as odd. These very odd and odd interviewers are under further monitoring. Here, in a second step we compare the proportion of the above-mentioned characteristic of NWRI and the proportion of those who are willing to participate in the second wave of the same interviewer (WRI). If the difference between the proportions also exceeds 10 percent here this interviewer number seems to be odd at second sight for this parameter.

All interviewer numbers with three characteristics described as very odd and odd are finally flagged as very problematic. This applies to five interviewers who conducted in total 163 interviews. If two of three characteristics (parenthood must be one of them) are identified as very odd and odd they are finally flagged as problematic. In total nine interviewers are flagged as very problematic or problematic. They conducted 2 to 3 percent of all interviews. By taking a closer look at the mean value of length of these interviews it can be shown that in most cases length is below average and that the time intervals between conducting different interviews is very short (often less than ten minutes).

If we now compare the results for childlessness or the share of single women to whether the problematic interviews are included or not, there are no extensive changes. Taking the share of childless western German women without problematic interviewers into consideration then in the older birth cohorts we have a minimal approximation of two percent points to the results of the Microcensus. Excluding problematic interviewers from the analysis therefore does not explain the total range of distortions, which is in line with results from other studies concerning the "quality" of fabricated interviews (Reuband 1990; Schnell 1991; Koch 1995; Schräpler and Wagner 2003).

We finally compared the problematic interviewers from the section on fertility history, partnership history, and the general interviewer check and identified four interviewers who were problematic in two out of three checks. Nevertheless we refrain from finally excluding them from the data as we only have strong indicators for problems with these interviewers, but cannot really prove if something went wrong and which interviews are affected.

Figure 4.2.6 Scheme of interviewer control



Own representation

5. Conclusion

The importance of collecting retrospective data that cover a long period of an individual's life is counterbalanced by concerns about the quality and reliability of such data. Although demography has been notable in the past for its attention to data quality, there were few recent evaluations of the quality of fertility or partnership histories in developed countries. As the GGS contains a great deal of retrospective data in order to cover a long period of individuals' lives, we were eager to validate this kind of data. Although landmark events such as the birth of children or marriage are commonly regarded as reliable memories and therefore are assumed to be applicable retrospectively without much decrement in data quality, we found severe distortions in the retrospective data on fertility and partnership in the German GGS of 2005 compared with the Microcensus. In fertility histories there is a great overestimation of childless women in older cohorts as well as an underestimation of this group in middle-aged cohorts. Regarding partnership history we have too many women who never married in our data in the older cohorts and too many married women in the younger cohorts.

We checked for possible sources of these distortions in the retrospective data in the GGS on several dimensions of the survey, starting with the event itself. Since births and living together with a partner are classified as very salient events in literature, we did not expect to find the cause of our problems there. We could not test for causes of distortion on the level of respondents' characteristics as missing or incorrect answers cannot be identified on an individual level and therefore cannot be assigned to individual respondents.

Our focus was then mainly directed at problems due to survey organisation. There we found indications of typical survey problems related to the random route sampling procedure of the GGS combined with an overly complex instrument and insufficient interviewer control. The complexity of the questionnaire may have been too demanding for respondents and/or interviewers. In any case, a result of this complexity was very time consuming interviews, which could have led to tiring effects as well as to conditioning and avoidance strategies concerning broad (follow-up) questions. We found evidence for this assumption in the so-called memory gap, which is larger in countries where control questions concerning the number of children were not implemented. We took the huge number of mothers who never married in the older cohorts as an indication of the correctness of this assumption. We considered the length of interviews a problem for interviewers as well as for respondents who would consequently try to shorten the time of the survey. Our focus, however, remained on interviewers as a weak point of the survey. We assumed that they should have an interest in quite short interviews as they were paid per interview and not according to duration. With every additional interview they also had a growing chance to learn how to best shorten interviews compared to respondents with only one interview. To test this assumption our focus was once again on the question of children living outside the household of the respondent as well as on the question about previous partnerships. In both cases the construction of the questionnaire was such that a simple "no" by the interviewer could shorten the interview substantially by omitting several loops. As there were no control mechanisms implemented in the questionnaire the risk of detection of such an omission would be quite small. An international comparison proved our hypothesis about the relevance of control questions. We could also show that several interviewers demonstrated "learning effects," i.e. the rank of interview had a significant effect on the probability of subjects having children outside the household or the probability of having had a previous partnership. We also tested for interviewer effects in several other ways and always found indications that there were incorrect codings, which affected the quality of the data although each one of them alone was too small to account for the whole extent of the distortions. All in all random route allows a great deal of uncontrolled interviewer

actions and the instrument lacks control mechanisms while simultaneously being very complex. We found problematic interviewers, but we could not delete the deviations in the data to our satisfaction as they are bound to be multicausal. We therefore conclude that there are combined effects of the sampling procedure and the complexity and length of the instrument that both add to interviewer effects.

This leads us to the question of how to proceed with the data of the German GGS. At this point we first should clarify that it was not only academic interest that led us to deal with event history data in the GGS, but in fact the problems that occurred when working with it. The section of the GGS that deals with the present situation of the respondents is not more affected than comparable surveys, as our comparison with the Microcensus has shown (see Chapter 3.1). It already worked with the data extensively and no comparable problems occurred there. Previous studies based on GGS data were confirmed by other surveys such as results on fertility intentions, which were confirmed by data of the Population Policy Acceptance Survey (Höhn, Ette, and Ruckdeschel 2006).

This leaves event history data in the German GGS as problematic, but even there not all of the data are equally affected. We could show that for the younger cohorts discrepancies with the Microcensus were mainly caused by random effects such as accessibility of respondents. That means that their representation may be biased. Another simple fact adds to the reliability of data on younger cohorts: the risk that they already have children outside the household or have lived together with many previous partners is quite small, because of their age. This leads to the fact that interviewers were not tempted so much to shorten interviews through miscoding. We therefore assume miscoding as a random, but not biasing effect. This changes when looking at older cohorts who are old enough to have children outside the household and had time enough to experience more than one partnership. Here we find indications of miscodings indeed. However we can narrow the group of contaminated data even further to those who did not indicate any past events, because those who have are not that critical. Thus the respondents in the older cohorts indicating no events are of our concern, because we do not know if the missing event is a real fact or an avoidance effect. One also must bear in mind that talking about a deviation of childless women of 11 percent points between the GGS and the Microcensus in cohort 1935-39 represents an absolute number of 29 women[21]. For a lot of research topics this may be no problem at all yet we nonetheless recommend using and interpreting the data of this group with great care. For research questions on fertility or partnership history of older cohorts we go as far as recommending not using them.

In any case, with our paper on the distortions in event history data in the German GGS and their possible reasons we wanted to provide a sound basis for every researcher to decide on his or her own whether or not the data are usable for their particular research questions. We did not check for all possible problems in such a large survey, so another aim of the paper is to remind colleagues to validate data before using them, preferably with official statistics or other surveys. Finally we wish to stress the point that a survey such as the GGS is no substitute for official vital statistics as there are always the usual survey problems that have to be taken into account. Nevertheless it provides rich and unique data for demographic research.

[21] 21% of 275 – 11% of 275, referring to western German women of German nationality

References

Aquilino, W.S. (1993). Effects of Spouse Presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly* 57(3): 358–376.

Auriat, N. (1991). Who forgets? An analysis of memory effects in a retrospective survey on migration history. *European Journal of Population/Revue européenne de Démographie* 7(4): 311–342. http://dx.doi.org/10.1007/BF01796872.

Babka v. Gostomski, C. and Hartmann, J. (1997). Zur Problematik von Retrospektivbefragungen. In: Kopp, J. (ed.). *Methodische Probleme der Familienforschung. Zu den praktischen Schwierigkeiten bei der Durchführung einer empirischen Untersuchung.* Frankfurt, New York: Campus: 113–146.

Backes, G.M. and Clemens, W. (1998). *Lebensphase Alter: Eine Einführung in die sozialwissenschaftliche Alternsforschung*. Weinheim und München: Juventa.

Beckett, M., DaVanzo, J., Sastry, N., Panis, C., and Peterson, C. (2001). The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country. *The Journal of Human Resources* 36(3): 595–625.

Belli, R.F. (1998). The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys. *Memory* 6(4): 383–406. http://www.informaworld.com/10.1080/741942610.

Belli, R.F., Shay, W.L., and Stafford, F.P. (2001). Event History Calendars and Conventional Questionnaire Surveys: A Direct Comparison of Interviewing Methods. *Public Opinion Quarterly* 65: 45–74.

Belli, R.F., Smith, L.M., Andreski, P.M., and Agrawal, S. (2007). Methodological Comparisons between CATI Event History Calendar and Standardized Conventional Questionnaire Instruments. *Public Opinion Quarterly* 71: 603–622.

Biemer, P.P. and Stokes, L. (1989). The Optimal Design of Quality Control Samples to Detect Interviewer Cheating. *Journal of Official Statistics* 5(1): 23–39.

Blossfeld, H.-P. and Huinink, J. (2001). Lebensverlaufsforschung als sozialwissenschaftliche Forschungsperspektive: Themen, Konzepte, Methoden und Probleme. *Zeitschrift für Biographieforschung und Oral History* 14(2): 5–31.

Bradburn, N.M. and Sudman, S. (1980). *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass Publishers.

Brüderl, J. and Klein, T. (2003). Die Pluralisierung partnerschaftlicher Lebensformen in Westdeutschland, 1960-2000: Eine empirische Untersuchung mit dem Familiensurvey 2000. In: Bien, W. and Marbach, J.H. (eds.). *Partnerschaft und Familiengründung. Ergebnisse der dritten Welle des Familien-Survey*. Opladen: Leske + Budrich: 189-217.

Buber, I. (2010). Parity-Specific Weights for the Austrian Generations and Gender Survey. Wien: Vienna Institute of Demography Working Papers 04/2010.

Bulgarian Academy of Sciences, Coordination Research Centre for Social Development and Social Eurointegration (2005). Final Report on the organization and conducting of the study GENERATION AND GENDER SURVEY after the execution of the working contract. Sofia: Max Planck Institute of Demographic Research.

Coombs, L.C. (1977). Levels of Reliability in Fertility Survey Data. *Studies in Family Planning* 8(9): 218–232. http://www.jstor.org/pss/1966189.

Crespi, L.P. (1945). The Cheater Problem in Polling. *Public Opinion Quarterly*: 431–445.

DeMaio, T.J. (1984). Social Desirability and Survey Measurement: A Review. In: Turner, C.F. and Martin, E. (eds.). *Surveying Subjective Phenomena 2*: New York: Russell Sage: 257–282.

Dorbritz, J. (2008). Germany: Family diversity with low actual and desired fertility: Childbearing Trends and Policies in Europe. Special Collection 7. *Demographic Research* 19: 557–598. http://www.demographic-research.org/Volumes/Vol19/17/19-17.pdf.

Dorbritz, J. and Ruckdeschel, K. (2007). Kinderlosigkeit in Deutschland - Ein europäischer Sonderweg? Daten, Trends und Gründe. In: Konietzka, D. and Kreyenfeld, M. (eds.). *Ein Leben ohne Kinder. Kinderlosigkeit in Deutschland*. Wiesbaden: VS Verlag: 45–81.

Duan, N., Alegria, M., Canino, G., McGuire, T.G., and Takeuchi, D. (2007). Survey Conditioning in Self-Reported Mental Health Service Use: Randomized Comparison of Alternative Instrument Formats. *Health Services Research* 42(2): 890–907.

Dykstra, P.A., Kalmijn, M., Knijn, T.C.M., Komter, A.E., Liefbroer, A.C., Mulder Clara H. (2005). Codebook of the Netherlands Kinship Panel Study, a multi-actor, multi-method panel study on solidarity in family relationships, Wave 1. The Hague: NKPS Working Paper No. 4.

Elder Jr., G.H. (2009). Perspectives on the Life Course. In: Heinz, W.R., Huinink, J., Weymann, A. (eds.). *The Life Course Reader. Individuals and Societies Across Time*. Frankfurt, New York: Campus: 91–110.

Esser, H. (1974). Der Befragte. In: van Koolwijk, J. and Wieken-Mayser, M. (eds.). *Techniken der empirischen Sozialforschung*, Bd. 4. München: Oldenburg: 107–145.

Festy, P. and Prioux, F. (2002). *An evaluation of the Fertility and Family Surveys project*. New York and Geneva: United Nations.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2009). *Survey Methodology*. 2nd Edition. Hoboken: Wiley.

Hartmann, P.H. and Schimpl-Neimanns, B. (1992). Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 44(2): 315–340.

Hatchett, S. and Schuman, H. (1975). White Respondents and Race-of-Interviewer Effects. *Public Opinion Quarterly* 39(4): 523–527.

Hayford, S.R. and Morgan, S.P. (2008). The Quality Of Retrospective Data On Cohabitation. *Demography* 45(1): 129–141.

Heyde, C. v. d. and Löffler, U. (1993). Die ADM-Stichprobe. *Planung und Analyse* 20(5): 49–53.

Hill, P.B. (2005). Methodenprobleme in der Ehe- und Familiensoziologie. In: Busch, F.W. and Nave-Herz, R. (eds.). *Familie und Gesellschaft. Beiträge zur Familienforschung*. Oldenburg: BIS-Verlag: 165–186.

Höhn, C., Ette, A., and Ruckdeschel, K. (2006). *Kinderwünsche in Deutschland - Konsequenzen für eine nachhaltige Familienpolitik*. Stuttgart: Robert-Bosch-Stiftung.

Holbrook, A.L., Green, M. C. and Krosnick, J.A. (2003). Telephone versus Face-to-face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly* 67: 79–125.

Jensen, P., Watanabe, H. K. and Richters, J. E. (1999). Who's Up First? Testing the Order Effects in Structured Interviews Using a Counterbalanced Experimental Design. *Journal of Abnormal Psychology* 27(6): 439–445.

Kane, E.W. and Macaulay, L.J. (1993). Interviewer Gender and Gender Attitudes. *Public Opinion Quarterly* 57: 1–28.

Karp, J.A. and Brockington, D. (2005). Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. *The Journal of Politics* 67(3): 825–840.

Khan, N. (2010). *An assessment of retrospective birth history reporting for the measurement of fertility in South Africa*. University of KwaZulu-Natal.

Klein, T. and Fischer-Kerli, D. (2000). Die Zuverlässigkeit retrospektiv erhobener Lebensverlaufsdaten: Analysen zur Partnerschaftsbiografie des Familiensurvey. *Zeitschrift für Soziologie* 29(4): 294–312.

Klijzing, E., Cairns, H. (2000). On the Quality of FFS Event History Data. Of critical Importance in Life Course Research. Bielefeld: Globalife Working Paper No. 13.

Knab, J.T. and McLanahan, S. (2007). Measuring Cohabitation: Does How, When, and Who you Ask Matter? In: Hofferth, S.L. and Casper, L.M. (eds.). *Handbook of measurment issues in family reasearch*. Mahwah, New Jersy: Lawrence Erlbaum Assoc Inc: 19–33.

Koch, A. (1995). Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994. *ZUMA-Nachrichten* 19(36): 89–105.

Koch, A. (1998). Wenn „mehr" nicht gleichbedeutend mit „besser" ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. *ZUMA-Nachrichten* 22(42): 66–93. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_42.pdf.

Kohler, U. (2007). Surveys from inside: An assessment of unit nonresponse bias with internal criteria. *Survey Research Methods* 1(2): 55–67. http://w4.ub.uni-konstanz.de/srm/article/view/75/53.

Kreuter, F., Mc Culloch, S., Presser, S., and Tourangeau, R. (2011). The Effects of Asking Filter Questions in Interleafed Versus Grouped Format. *Sociological Methods & Research* 40(1): 88–104.

Kreyenfeld, M. and Konietzka, D. (2008). Education and fertility in Germany. In: Hamm, I., Seitz, H., Werding, M. (eds.). *Demographic change in Germany: the economic and fiscal consequences*. Berlin, Heidelberg, New York: Springer: 164–187.

Kreyenfeld, M., Hornung, A., Kubisch, K., Jaschinski, I. (2010). Fertility and union histories from German GGS data: some critical reflections. Rostock: Max Planck Institute for Demographic Research Working Paper 2010-023.

Kreyenfeld, M., Konietzka, D., and Walke, R. (2011). Dynamik und Determinanten nichtehelicher Mutterschaft in Ost- und Westdeutschland. In: Brüderl, J., Castiglioni, L., Schumann, N. (eds.). *Partnerschaft, Fertilität und intergenerationale Beziehungen. Ergebnisse der ersten Welle des Beziehungs- und Familienpanels*. Würzburg: Ergon Verlag: 155–174.

Kreyenfeld, M., Zeman, K., Burkimsher, M., Jaschinski, I. (2011). Fertility data for German-speaking countries: What is the potential? Where are the pitfalls? Rostock: Max Planck Institute for Demographic Research Working Paper 2011-003.

Krosnick, J.A. (1991). Response strategies for coping with the cognitive demands of attitudes measures in surveys. *Applied Cognitive Psychology* 5(3): 213–236.

Krosnick, J.A. and Presser, S. (2010). Question and Questionnaire Design. In: Marsden, P.V. and Wright, J.D. (eds.). *Handbook of Survey Research*. 2nd edition. Bingley, UK: Emerald: 263–314.

Lappegård, T., Veenstra, M. (2010). *Life-course, generation and gender. LOGG 2007: Field report of the Norwegian Generations and Gender Survey Statistisk*. Oslo–Kongsvinger.

Loosveldt, G., Carton, A., and Pickery, J. (2008). The Effect of Interviewer and Respondent Characteristics on Refusals in a Panel Survey. In: Koch, A. and Porst, R. (eds.). *Nonresponse in Survey Research*. *Proceedings of the Eighth International Workshop on Household Survey Nonresponse 24-26 Sept*. 1997. Mannheim: 249–262.

Lucas, C.P., Fisher, P., Piacentini J., Zhang H., Jensen P.S., Shaffer D., and Dulcan M., Schwab-Stone M., Regier D., and Canino G. (1999). Features of Interview Questions Associated with Attenuation of Symptom Reports. *Journal of Abnormal Psychology* 27(6): 429–437.

Marsden, P.V. (2003). Interviewer Effects in Measuring Network Size Using a Single Name Generator. *Social Networks* 24: 407–422.

Mayer, K.U. (2009). New Directions in Life Course Research. *Annual Review of Sociology* (35): 413–433.

Mitchell, C. (2010). Are divorce studies trustworthy? The effects of survey nonresponse and response errors. *Journal of Marriage and Family* 72(4): 893-095.

Murphy, M. (2000). The Evaluation of Cohabitation in Britain, 1960-95. *Population Studies* 54(1): 43–56. http://www.jstor.org/stable/2584632.

Murphy, M. (2009). Where have all the children gone? Women‘s reports of more childlessness at older ages than when they were younger in a large-scale continuous household survey in Britain. *Population Studies* 63(2): 115–133.

Neller, K. (2005). Kooperation und Verweigerung: Eine Non-Response-Studie. *ZUMA-Nachrichten* 29(57): 9–36. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_57.pdf.

Papastefanou, G. and Tölke, A. (1981). Zur Adäquanz retrospektiver Daten. In: Schulte, W. (ed.). *Soziologie in der Gesellschaft: Referate aus den Veranstaltungen der Sektionen der Deutschen Gesellschaft für Soziologie, der Ad-hoc-Gruppen und des Berufsverbandes Deutscher Soziologen beim 20. Deutschen Soziologentag in Bremen 1980*. Bremen: 629–634.

Peters, H.E. (1988). Retrospective Versus Panel Data in Analyzing Lifecycle Events. *The Journal of Human Resources* 23(4): 488–513.

Pötzsch, O. (2010). Cohort Fertility: A Comparison of the Results of the Official Birth Statistics and of the Microcensus Survey 2008. *Comparative Population Studies* 35(1): 185–204.

Pötzsch, O. and Emmerling, D. (2008). *Geburten und Kinderlosigkeit in Deutschland: Bericht über die Sondererhebung 2006 „Geburten in Deutschland“*. Wiesbaden: Statistisches Bundesamt.

Poulain, M., Riandey, B., and Firdion, J.-M. (1992). Data from a life history survey and from Belgian Population Register: A comparison. *Population: An English Selection* 4: 77–96.

Regnier-Loilier, A. (2006). *Présentation, questionnaire et documentation de l’« Etude des relations familiales et intergénérationnelles » (ERFI): Version française de l’enquête « Generations and Gender Survey » (GGS)*. Paris: Documents de travail 133.

Reimondos, A., Evans, A. and Gray, E. (2011). Reports of relationship timing: Missing data and couple agreement. *Survey Research Methods* 5(5): 75–87.

Reuband, K.-H. (1990). Interviews, die keine sind - „Erfolge“ und „Mißerfolge“ beim Fälschen von Interviews. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (4): 706–733.

Ross, M. and Holmberg, D. (1992). Are wives‘ memories for events in relationships more vivid than their husband‘s memories. *Journal of Social and Personal Relationships* 9: 585-604.

Ruckdeschel, K., Ette, A., Hullen, G., Leven, I. (2006). *Generations and Gender Survey: Dokumentation der ersten Welle der Hauptbefragung in Deutschland*. Wiesbaden: Materialien zur Bevölkerungswissenschaft 121.

Schaeffer, N.C. and Presser, S. (2003). The Science of Asking Questions. *Annual Review of Sociology* 29: 65–88.

Schaeffer, N.C., Dykema, J., and Maynard, D.W. (2010). Interviews and Interviewing. In: Marsden, P.V. and Wright, J.D. (eds.). *Handbook of Survey Research. 2nd edition*. Bingley, UK: Emerald: 437–470.

Schmidt, G., Matthiesen, S., Dekker, A., and Starke, K. (2006). *Spätmoderne Beziehungswelten*. Wiesbaden: VS Verlag.

Schnell, R. (1991). Der Einfluß gefälschter Interviews auf Survey-Ergebnisse. *Zeitschrift für Soziologie* 20(1): 25–35.

Schnell, R. (1997). *Nonresponse in Bevölkerungsumfragen: Ausmaß, Entwicklung und Ursachen*. Opladen: Leske+Budrich.

Schräpler, J.-P. and Wagner, G.G. (2003). Identification, Characteristics and Impact of Faked Interviews in Surveys: An Analysis by means of genuine fakes in raw data of SOEP. Bonn: IZA Discussion Paper No. 969.

Schreiner, I., Pennie, K., and Newbrough, J. (1988). Interviewer Falsification in Census Bureau Surveys. Proceedings of the American Statistical Association. *Survey Research Methods Section*: 491–496.

Silver, B.D., Anderson, B.A., and Abramson, P.R. (1986). Who overreports voting? *American Political Science Review* 80(2): 613–624.

Singh, S. (1987). Evaluation of Data Quality. In: Cleland, J.G., Scott, C., Whitelegge, D. (eds.). *The World Fertility Survey: An assessment*. Oxford, New York: Published for the International Statistical Institute by Oxford University Press: 618–643.

Skowronski, J.J. and Thompson, C.P. (1990). Reconstructing the dates of personal events: Gender differences in accuracy. *Applied Cognitive Psychology* 4: 371–381.

Skowronski, J.J., Betz, A.L., Thompson, C.P., Walker, W.R., and Shannon, L. (1994). The Impact of Differing Memory Domains on Event Dating Processes in Self and Proxy Preports. In: Schwarz, N. and Sudman, S. (eds.). *Autobiographical Memory and the Validity of Retrospective Reports*. New York, Berlin, Heidelberg: Springer-Verlag: 217–231.

Smith, J.P. and Thomas, D. (2003). *Remembrances of things past: test–retest reliability of retrospective migration histories*. *Journal of the Royal Statistical Society: Series A (Statistics in Society*) 166(1): 23–49. http://dx.doi.org/10.1111/1467-985X.00257.

Sodeur, W. (2007). Entscheidungsspielräume von Interviewern bei der Wahrscheinlichkeitsauswahl: Ein Vergleich von ALLBUS-Erhebungen. *Methoden – Daten – Analysen* 1(2): 107–130. http://www.ssoar.info/fileadmin/php/download.php?url=/ssoar/files/2010/1398/MDA_2007_2_sodeur.pdf.

Statistik Austria (2009). *Generations and Gender Survey: Familienentwicklung in Österreich Welle 1* - Generationen und Geschlechterrollen. Wien: Endbericht.

Statistisches Bundesamt (2009a). Imputation von Werten bei fehlenden Angaben zur Mutterschaft und zur Zahl der geborenen Kinder im Mikrozensus 2008: Arbeitsbericht. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt (2009b). Mikrozensus 2008: Neue Daten zur Kinderlosigkeit in Deutschland: Begleitmaterial zur Pressekonferenz am 29. Juli 2009 in Berlin. Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt (2011). *Bevölkerung und Erwerbstätigkeit: Natürliche Bevölkerungsbewegung 2009*. Fachserie 1 Reihe 1.1. Wiesbaden: Statistisches Bundesamt.

Sudman, S. and Bradburn, N.M. (1974). *Response Effects in Surveys. A Review and Synthesis*. Chicago: Aldine.

Sudman, S., Bradburn, N.M., and Schwarz, N. (1996). *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francicso: Jossey-Bass Publishers.

Swicegood, C., Gray, S., Morgan, P., and Rindfuss, R.R. (1984). Measurement and Replication: Evaluating the Consistency of Eight U.S. Fertility Surveys. *Demography* 21: 15.

Teitler, J.O., Reichman, N.E., and Koball, H. (2006). Contemporaneous Versus Retrospective Reports of Cohabitation in the Fragile Families Survey. *Journal of Marriage and Family* 68(2): 469–477. http://dx.doi.org/10.1111/j.1741-3737.2006.00265.x.

Tourangeau, R. and Smith, T.W. (1996). Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly* 60: 275–304.

Tourangeau, R. and Smith, T.W. (1998). Collecting sensitive information with different modes of data collection. In: Couper, M., R. Baker, R., Bethlehem, J., Clark, C., Martin, C., Nicholls, W. et al. (eds.). *Computer Assisted Survey Information Collection*. New York: Wiley & Sons: 431–454.

Tourangeau, R., Rips, L.J., and Rasinski, K. (2000). *The Psychology of Survey Response*. *Cambridge*: Cambridge University Press.

Traugott, M.W. and Katosh, J.P. (1979). Response Validity in Surveys of Voting Behaviour. *Public Opinion Quarterly* 43: 359–377.

United Nations (ed.) (1987). *A Comparative evaluation of data quality in thirty-eight world fertility surveys*. New York: United Nations.

van Tilburg, T. (1998). Interviewer effects in the measurement of personal network size. *Sociological Methods & Research* 26: 300–328.

Vikat, A., Spéder, Z., Beets, G., Billari, F.C., Bühler, C., Désesquelles, A., Fokkema, T., Hoem, J.M., MacDonald, A., Neyer, G., Pailhé, A., Pinnelli, A., Solaz, A. (2007). Generations and Gender Survey (GGS): Towards a better understanding of relationships and processes in the life course. *Demographic Research* 17(14): 389–440. http://www.demographic-research.org/volumes/vol17/default.htm.

Wagenaar, W.A. (1986). My memory: A Study of Autobiographical Memory over Six Years. *Cognitive Psychology* 18(1): 225–252.

Winkler-Dworak, M. and Engelhardt, H. (2004). On the tempo and quantum of first marriages in Austria, Germany, and Switzerland: Changes in mean age and variance. *Demographic Research* 10: 232–263.

Wolf, C. (2004). Egozentrierte Netzwerke. Erhebungsverfahren und Datenqualität. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 44: 244–273.

Wu, L.L., Martin, S.P., and Long, D.A. (2001). Comparing Data Quality of Fertility and First Sexual Intercourse Histories. *Journal of Human Resources* 36: 520–555.

Appendix

Table A1: Distribution of characteristics in the German GGS and Microcensus 2005

	GGS		MC 2005*	
Composition of population				
German nationals	94.53		89.82	
Foreign nationals	5.47		10.18	
N	10,000		64,787,480	
Number of household members(only German nationals)				
One-person household	25.58		34.95	
Two-person household	35.57		35.58	
Three-person household	17.32		14.65	
Four-person household	15.38		11.16	
More than four-person household	6.14		3.66	
N	9,451		32,773,390	
Sex (only German nationals)				
Male	45.91		49.15	
Female	54.09		50.85	
N	9,451		58,195,340	
Age distribution (only German nationals)	**Male**	**Female**	**Male**	**Female**
1925-1929	4.68	4.86	4.00	5.68
1930-1934	6.00	5.69	5.47	6.47
1935-1939	8.98	7.38	8.26	8.98
1940-1944	9.98	7.65	8.24	8.49
1945-1949	7.76	6.79	7.02	6.98
1950-1954	8.78	8.70	8.72	8.61
1955-1959	9.22	10.70	9.52	9.26
1960-1964	11.28	12.20	11.31	10.77
1965-1969	9.33	11.76	10.61	9.98
1970-1974	6.02	8.44	7.61	7.18
1975-1979	6.74	6.69	6.89	6.41
1980-1984	7.39	6.33	7.67	7.01
1985-1987	3.87	2.81	4.68	4.19
N	4,319	5,083	28,601,980	29,593,130
Educational level**		**German nationals**		
	Male	**Female**	**Male**	**Female**
Still in school	2.33	2.96	2.02	2.05
Primary and lower secondary education (*German Haupt-/ Volksschulschulabschluss*)	42.41	34.44	41.37	35.11
Lower secondary education (*German Realschulabschluss or Abschluss einer Polytechnischen Oberschule*)	16.73	26.30	16.05	16.69
Upper secondary education (*German Fachhochschulreife or Abitur*)	31.52	31.85	26.26	26.64
Left school without diploma	7.00	4.44	14.30	19.51
N	257	270	3,321,000	3,127,000

Notes: * Private households. Main and secondary residence. Cohorts 1925-87
** For reasons of proper classification we used the original German data set to compute educational level

Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

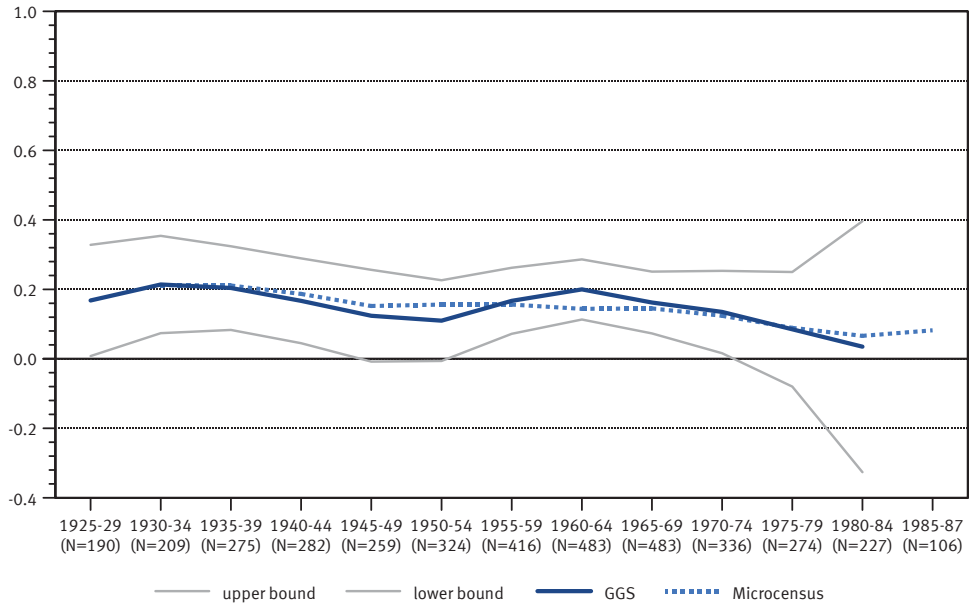
Imputation procedure for missing values on the question about children ever born in the Microcensus 2008

To correct the bias caused by missing answers on the question about children ever born in the Microcensus 2008 a two-step imputation procedure was used. In a first step, it was taken into account if any non-married children were living in the household who indicated the respondent as their mother. In this case the respondent was coded as mother. The risk that these persons were adoptive children, stepchildren or foster children was estimated as about 0.6 percent[22] which is negligibly small. With this first step 29 percent of all missing values were substituted. It was trickier to do this for women in households without children. The data of those women who answered the question about children ever born showed that age, marital status, and education were crucial factors for estimating if a woman ever had children. With this information the probability for childlessness was estimated for women with non-response. If the probability was higher than 90 percent, the missing value was replaced by the indication "childless" in all other cases the missing value remained. With this procedure another 30 percent of the non-responses were replaced (see Statistisches Bundesamt 2009a). In a second step the number of children was estimated with the help of available information. For this the probability that the number of children in the household equals the total number of children was calculated. Apart from age and educational level this time also regional factors proved to be important, because children in eastern Germany move out of their parents' home at a younger age than in western Germany. In total 10 percent of missing values were replaced in this second step.

This imputation procedure was quite successful for younger cohorts. Thus, for example, the non-response rate for women aged 30 could be reduced from 14 percent to 3 percent. In contrast, in the cohorts 1950 and older, there still remains a relatively high non-response rate after the imputation.

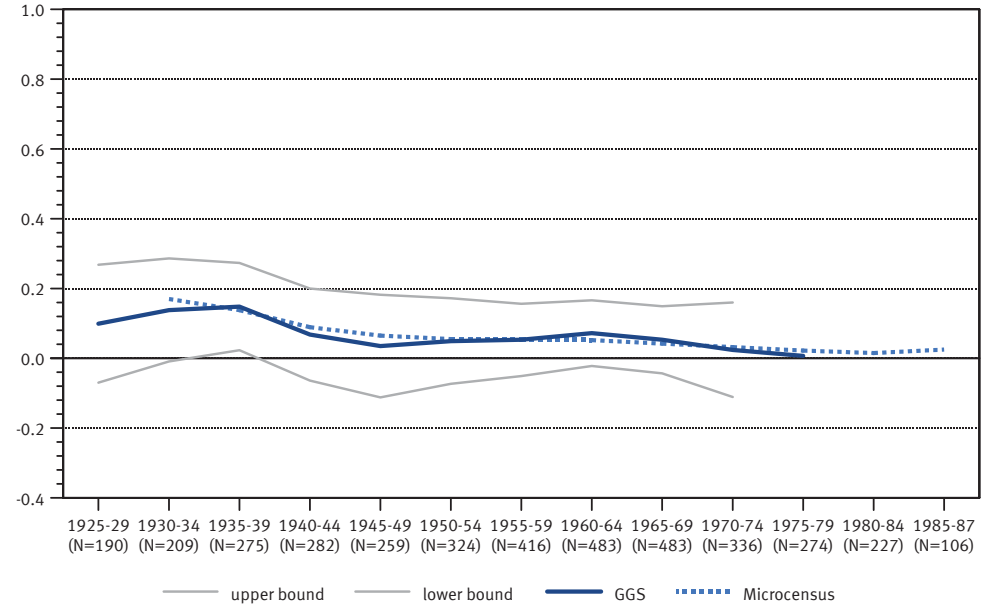
[22] Risk calculated with data of women who answered the question.

Figure A.1: Proportion of mothers with parity 3 of all mothers by cohort, Microcensus 2008 and GGS 2005



Notes: Women of German nationality living in western Germany
Microcensus cohorts starting with 1933
N=all members of selected cohorts in GGS
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

Figure A.2: Proportion of mothers with parity 4+ of all mothers by cohort, Microcensus 2008 and GGS 2005



Notes: Women of German nationality living in western Germany
Microcensus cohorts starting with 1933
N=all members of selected cohorts in GGS
Data sources: German GGS V 3.0 unweighted data; Microcensus by remote execution, own calculations

Table A2: Percentage and absolute numbers for Figure 4.2.4 on the proportion of mothers among all married and among all women who never married by cohort, GGS 2005

Cohorts	Never married		Married	
	%	N	%	N
1925-29	*63*	127	*78*	59
1930-34	*71*	125	*81*	101
1935-39	*71*	111	*83*	188
1940-44	*73*	75	*83*	234
1945-49	*73*	67	*80*	222
1950-54	*65*	94	*87*	294
1955-59	*71*	114	*90*	372
1960-64	*71*	166	*93*	416
1965-69	*61*	159	*93*	404
1970-74	*53*	156	*88*	259
1975-79	*29*	195	*84*	134
1980-84	*19*	273	*74*	42
1985-87	*7*	138	*33*	3

Notes: Women of German nationality living in western Germany
N=all members of selected group
Data sources: German GGS V 3.0 unweighted data, own calculations

Table A3: Laptop/interviewer number and willingness to be re-contacted

Laptop number (i)	Number of realised interviews	Number of respondents who did not agree to be re-contacted	Willingness of being re-contacted in %
i=1	130	130	0
i=2	80	80	0
i=3	50	50	0
i=4	40	40	0
i=5	36	36	0
i=6	34	34	0
i=7	33	33	0
i=8	31	31	0
i=9	30	30	0
i=10	29	29	0
i=11	29	29	0
i=12	28	28	0
i=13	27	27	0
i=14	26	26	0
i=15	24	24	0
i=16	22	22	0
i=17	20	20	0
i=18	19	19	0
i=19	18	18	0
i=20	17	17	0
i=21	16	16	0
i=22	15	15	0
i=23	14	14	0
i=24	12	12	0
i=25	12	12	0
i=26	11	11	0
i=27	11	11	0
i=28	10	10	0
i=29	10	10	0
i=30	10	10	0
i=31	10	10	0
i=32	10	10	0
i=33	90	89	1.11
i=34	50	49	2.00
i=35	40	39	2.50
i=36	24	23	4.17
i=37	22	21	4.55
i=38	15	14	6.67
i=39	28	26	7.14
i=40	27	25	7.41
i=41	13	12	7.69
i=42	61	56	8.20
I=43	48	44	8.33
i=44	12	11	8.33
i=45	23	21	8.70
i=46	11	10	9.09
i=47	21	19	9.52
i=48	31	28	9.68
i=49	10	9	10.00
Total	**1,390**	**1,360**	
25 laptop numbers	**645**	**478**	**11-40%**
30 laptop numbers	**658**	**321**	**41-60%**

Own representation